

THE SEMICLASSICAL MODIFIED NONLINEAR SCHRÖDINGER EQUATION II: ASYMPTOTIC ANALYSIS OF THE CAUCHY PROBLEM. THE ELLIPTIC REGION FOR TRANSSONIC INITIAL DATA.

JEFFERY C. DIFRANCO AND PETER D. MILLER

ABSTRACT. We begin a study of a multi-parameter family of Cauchy initial-value problems for the modified nonlinear Schrödinger equation, analyzing the solution in the semiclassical limit. We use the inverse scattering transform for this equation, along with the steepest descent method of Deift and Zhou. The initial conditions are selected both to allow all relevant scattering data to be calculated without approximation and also to place the governing equation in a transsonic state in which the quantum fluid dynamical system formally approximating it is of hyperbolic type for some x and of elliptic type for other x . Our main result is a global approximation theorem valid in a maximal space-time region connected to the elliptic part of the initial data.

1. INTRODUCTION

Let ϵ and α be positive parameters. The modified nonlinear Schrödinger (MNLS) equation

$$(1) \quad i\epsilon \frac{\partial \phi}{\partial t} + \frac{\epsilon^2}{2} \frac{\partial^2 \phi}{\partial x^2} + |\phi|^2 \phi + i\alpha \epsilon \frac{\partial}{\partial x} (|\phi|^2 \phi) = 0, \quad x \in \mathbb{R}, \quad t > 0$$

is a completely integrable generalization of the focusing nonlinear Schrödinger equation, to which the MNLS equation reduces upon setting $\alpha = 0$. The perturbation term proportional to α models the effect of *nonlinear dispersion* and is one of a suite of three terms arising in small-amplitude perturbation theory one order beyond the focusing nonlinear Schrödinger equation in the analysis of short pulse propagation in weakly nonlinear optical fibers [11, 12, 13]. Although the MNLS equation appears to be a perturbation of the focusing nonlinear Schrödinger equation, there is a sense in which it can also be considered to be a perturbation of the defocusing nonlinear Schrödinger equation. This is related to broken Galilean symmetry introduced for $\alpha \neq 0$, which also shows that the MNLS equation can be reduced to the so-called derivative nonlinear Schrödinger equation of Kaup and Newell [15]. The complete integrability of the MNLS equation is connected with its representation as the compatibility condition of a Lax pair based on the WKI spectral problem of Wadati, Konno, and Ichikawa [22]. The existence of a Lax pair representation makes it possible to analyze quite general solutions of the MNLS equation with remarkable accuracy; as an example, see the long-time asymptotic analysis of Kitaev and Vartanian [16, 17].

The Cauchy initial-value problem for (1) is to find a solution ϕ subject to the initial condition:

$$(2) \quad \phi(x, 0) = A_0(x) e^{iS_0(x)/\epsilon}, \quad x \in \mathbb{R}.$$

Here, $A_0(\cdot)$ and $S_0(\cdot)$ are real-valued amplitude and phase functions respectively. Under suitable conditions on these two functions, the Cauchy initial-value problem can be studied with the help of an *inverse-scattering transform* derived from Lax pair representation of the MNLS equation (1).

Our main interest is in the asymptotic behavior of $\phi = \phi_\epsilon(x, t)$ in the *semiclassical limit* where $\epsilon \downarrow 0$ with $\alpha > 0$ and the functions $A_0(\cdot)$ and $S_0(\cdot)$ held fixed. This limit is obviously very singular, but the problem can be recast in a form that appears somewhat more tractable by introducing Madelung's fluid dynamical variables [19]:

$$(3) \quad \begin{aligned} \rho_\epsilon(x, t) &:= |\phi_\epsilon(x, t)|^2, & (\text{a quantum fluid density}) \\ u_\epsilon(x, t) &:= \epsilon \Im \left\{ \frac{\partial}{\partial x} \log(\phi_\epsilon(x, t)) \right\}, & (\text{a quantum fluid velocity}). \end{aligned}$$

This work was partially supported by the National Science Foundation under grant numbers DMS-0807653 and DMS-1206131.

Here of course $\log(\phi_\epsilon(x, t))$ is defined (assuming $\rho_\epsilon(x, t)$ is non-vanishing) by continuation to be a differentiable function of x . The initial values of these fields are independent of ϵ , since according to (2),

$$(4) \quad \rho_\epsilon(x, 0) = \rho_0(x) := A_0(x)^2 \quad \text{and} \quad u_\epsilon(x, 0) = u_0(x) := S'_0(x).$$

Most of what we will discuss in this paper has to do with quite special choices for the functions $A_0(\cdot)$ and $S_0(\cdot)$. However, to ensure the validity of certain general asymptotic formulae for scattering data (see (46) and (61)), we assume that the functions $\rho_0(\cdot)$ and $u'_0(\cdot)$ are Schwartz-class real-analytic functions of $x \in \mathbb{R}$, and that $\rho_0(\cdot)$ is nowhere vanishing. In particular, these assumptions imply that $u_0(x) \rightarrow u_\pm$ as $x \rightarrow \pm\infty$. Then

$$(5) \quad S_0(x) = S_0(0) + \int_0^x u_0(y) dy \quad \implies \quad S_0(x) = u_\pm x + S_\pm + o(1), \quad x \rightarrow \pm\infty,$$

where

$$(6) \quad S_+ := S_0(0) + \int_0^{+\infty} [u_0(y) - u_+] dy \quad \text{and} \quad S_- := S_0(0) - \int_{-\infty}^0 [u_0(y) - u_-] dy.$$

In terms of the Madelung fields defined by (3), the MNLS equation (1) can be rewritten without approximation in the form of a coupled system:

$$(7) \quad \frac{\partial \rho_\epsilon}{\partial t} + \frac{\partial}{\partial x} \left(\rho_\epsilon u_\epsilon + \frac{3}{2} \alpha \rho_\epsilon^2 \right) = 0 \quad \text{and} \quad \frac{\partial u_\epsilon}{\partial t} + \frac{\partial}{\partial x} \left(\frac{1}{2} u_\epsilon^2 - \rho_\epsilon + \alpha \rho_\epsilon u_\epsilon \right) = \frac{1}{2} \epsilon^2 \frac{\partial F[\rho_\epsilon]}{\partial x},$$

where $F[\rho_\epsilon]$ is given by the differential rational expression

$$(8) \quad F[\rho_\epsilon] := \frac{1}{2\rho_\epsilon} \frac{\partial^2 \rho_\epsilon}{\partial x^2} - \left(\frac{1}{2\rho_\epsilon} \frac{\partial \rho_\epsilon}{\partial x} \right)^2.$$

The only explicit dependence on ϵ in either the equations of motion (7) or the initial conditions (4) lies in the dispersive correction term on the right-hand side of the equation governing u_ϵ in (7), and it seems reasonable to try to neglect that term and solve the corresponding quasilinear system of local conservation laws, the *dispersionless MNLS system*

$$(9) \quad \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} \left(\rho u + \frac{3}{2} \alpha \rho^2 \right) = 0 \quad \text{and} \quad \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{1}{2} u^2 - \rho + \alpha \rho u \right) = 0,$$

with the ϵ -independent initial data $\rho(x, 0) = \rho_0(x)$ and $u(x, 0) = u_0(x)$ given by (4). Although this seems attractive, proving that such a procedure yields an accurate approximation when $\epsilon \ll 1$ is not at all straightforward for several reasons. Perhaps the greatest obstruction lies in the fact that the dispersionless MNLS system (9) is not strictly hyperbolic, leading to a certain ill-posedness of the ϵ -independent Cauchy initial value problem purported to approximate the true dynamics. This means that without strong assumptions on the initial data, there may exist no corresponding solution of (9) at all. Even if there is a solution, linearization about a constant state reveals unbounded exponential growth rates corresponding to modes that are likely to be seeded by the dispersive correction term (8); it is not obvious at all in such a situation whether the solution for small nonzero ϵ should resemble for t strictly positive that obtained by simply setting $\epsilon = 0$.

The characteristic velocities of the dispersionless MNLS system (9) satisfy a quadratic equation with real coefficients depending on the local values of the Madelung variables ρ and u . The discriminant of this quadratic is proportional via a positive factor to

$$(10) \quad Q := \alpha^2 \rho + \alpha u - 1.$$

At any given $t \geq 0$ the fields $\rho(\cdot, t)$ and $u(\cdot, t)$ can cause the system to be in any one of three different states (with terminology borrowed from the language of stationary flows in gas dynamics):

- *Globally supersonic.* In this case the dispersionless MNLS system is strictly hyperbolic, corresponding to $Q > 0$ and hence real distinct characteristic velocities, for all $x \in \mathbb{R}$.
- *Globally subsonic.* In this case, the dispersionless MNLS system is elliptic, corresponding to $Q < 0$ and hence distinct complex-conjugate characteristic velocities, for all $x \in \mathbb{R}$.
- *Transsonic.* In this case, the dispersionless MNLS system is hyperbolic for some values of x and elliptic for others, and Q changes sign as a function of x .

While the correspondence is not the obvious one, it is shown in [10] that for initial data (4) that makes the dispersionless MNLS system (9) globally supersonic (respectively subsonic) at $t = 0$, the semiclassical analysis of the MNLS Cauchy problem can essentially be reduced to that of the defocusing (respectively focusing) cubic nonlinear Schrödinger equation. This result generalizes to arbitrary genus a fact that was found for genus zero in [8] and for genus one in [18].

Evidently only for transsonic initial data might the MNLS equation (1) behave unlike either the focusing or defocusing cubic nonlinear Schrödinger equation in the semiclassical limit. *This observation is the key motivation for the study that we begin in this paper.*

To analyze the solution of the MNLS equation (1) subject to suitable initial data of the form (2) by means of the inverse-scattering transform, one must in general study the small- ϵ asymptotics of two types of problems:

- The direct scattering problem. This amounts to the asymptotic analysis of a singularly perturbed linear differential equation (the WKI spectral problem). The natural tools applicable to this problem include the classical WKB method and its generalizations.
- The inverse scattering problem. This can be formulated as a matrix Riemann-Hilbert problem (see §3) with a highly oscillatory or rapidly exponentially growing and decaying jump matrix. The natural tools applicable to this problem include the noncommutative steepest descent method of Deift and Zhou [7] and its generalizations (in particular the mechanism of the so-called *g-function* first introduced in [6]).

Although the asymptotic analysis of the direct scattering problem can be carried out in part using classical methods, what really makes the problem difficult is (i) that one requires asymptotics that are uniform with respect to the spectral parameter, including near exceptional values where turning points collide, and (ii) that for some initial data it is required to approximate some exponentially small quantities that — like the semiclassical above-barrier reflection coefficient for the Schrödinger operator — cannot be suitably estimated without assumptions of analyticity and the use of turning point analysis in the complex x -plane. For a flavor of the type of calculations required to analyze the WKI direct scattering problem for the MNLS equation in the semiclassical limit, see [10]. It should be stressed however, that with very few exceptions in the literature¹ the asymptotic information that is available from the direct problem is frequently of insufficient quality and/or accuracy to continue with the inverse scattering problem without unjustified formal approximations. This is especially true when (as for the WKI problem or the Zakharov-Shabat problem for the focusing nonlinear Schrödinger equation) the direct spectral problem cannot be cast as an eigenvalue problem for a self-adjoint operator.

On the other hand, the completely rigorous asymptotic analysis of matrix Riemann-Hilbert problems is now a rather well-developed science. In order to bypass the difficulties mentioned above in regard to the direct problem and hence enable rigorous analysis of the Cauchy problem for the MNLS equation (1), we will choose the initial condition functions $A_0(x)$ and $S_0(x)$ to be of the specific form

$$(11) \quad A_0(x) = \nu \operatorname{sech}(x) \quad \text{and} \quad S_0(x) = S_0(0) + \delta x + \mu \log(\cosh(x))$$

for real parameters $\nu \neq 0$, $S_0(0)$, δ , and μ . The corresponding initial values for the Madelung fields are

$$(12) \quad \rho_0(x) = \nu^2 \operatorname{sech}^2(x) \quad \text{and} \quad u_0(x) = \delta + \mu \tanh(x).$$

In fact, without loss of generality, we will assume that $\nu = 1$ and $S_0(0) = 0$, as this can be accomplished by making the substitutions

$$(13) \quad \phi = \nu e^{iS_0(0)/\epsilon} \tilde{\phi}, \quad \epsilon = |\nu| \tilde{\epsilon}, \quad \alpha = |\nu|^{-1} \tilde{\alpha}, \quad t = |\nu| \tilde{t}$$

(x is not scaled) in both (1) and (11) and dropping the tildes. Unlike the cubic nonlinear Schrödinger equation, the MNLS equation is not invariant under Galilean boosts [8] so the parameter δ cannot be removed even by going to a moving frame of reference. Therefore, both δ and μ are essential parameters in the family of initial data of the form (11). The advantage of working with the family (11) is that this choice

¹A notable exception is the analysis of the Korteweg-de Vries equation in the small dispersion limit by Claeys and Grava [3, 4, 5]. Of course in this case the direct scattering problem corresponds to the well-studied and self-adjoint Schrödinger operator, and sufficiently accurate semiclassical approximations of the scattering data for this problem were obtained (only fairly recently) by Ramond [20].

renders the direct scattering problem for the MNLS equation equivalent for all $\epsilon > 0$ to a hypergeometric equation, and this in turn allows the corresponding scattering data to be obtained without any approximation whatsoever [8].

For the initial data (12) with $\nu = 1$, the quantity Q defined by (10) becomes (for $t = 0$)

$$(14) \quad Q(x) = \alpha^2 \operatorname{sech}^2(x) + \alpha\delta + \alpha\mu \tanh(x) - 1 = -\alpha^2 T^2 + \alpha\mu T + \alpha^2 + \alpha\delta - 1, \quad T := \tanh(x).$$

To be in the transsonic case, we need $Q(x)$ to change sign. As a quadratic function of $T = \tanh(x)$ it is clear that there are at most two roots. For simplicity we want to arrange that there is exactly one simple root of this quadratic in the interval $T \in (-1, 1)$, and this will occur if and only if it takes opposite signs for $T = \pm 1$; hence we assume the condition

$$(15) \quad \alpha|\mu| > |1 - \alpha\delta|.$$

For technical reasons that we will explain in §2.2 we also assume the following two conditions:

$$(16) \quad \mu^2 > 4(1 - \alpha\delta),$$

and

$$(17) \quad \mu > 0.$$

If we introduce the combined parameters

$$(18) \quad A := \frac{1 - \alpha\delta}{4\alpha^2} \quad \text{and} \quad B := \frac{\mu}{4\alpha},$$

then conditions (15)–(17) imply that (A, B) should occupy the shaded region in the diagram pictured in Figure 1. As an example, some specific parameter values that satisfy these conditions are $\alpha = \delta = 1$ and

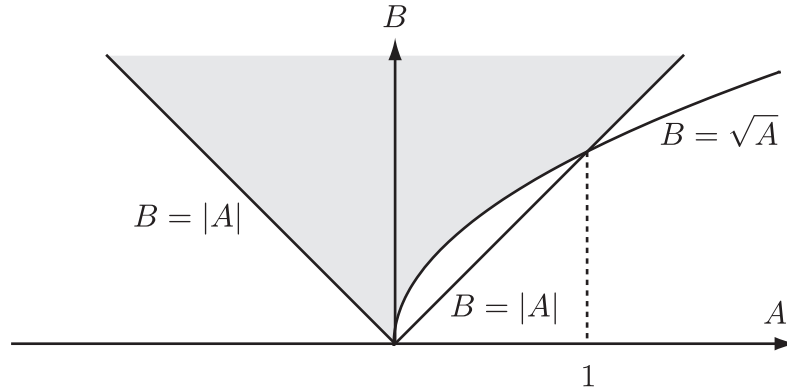


FIGURE 1. The shaded region corresponds to the inequalities (15)–(17).

$\mu = 2$, corresponding to $A = 0$ and $B = 1/2$. Under the inequalities (15)–(17), the unique simple root in the interval $(-1, 1)$ of the quadratic (14) is exactly

$$(19) \quad T = T_c := 2B - \sqrt{4B^2 - 4A + 1}$$

(the other root exceeds $T = 1$ for (A, B) in the admissible region), and it follows that when $t = 0$, $Q < 0$ (elliptic case) for $x < x_c$ and $Q > 0$ (hyperbolic case) for $x > x_c$, where $x_c := \operatorname{arctanh}(T_c)$. Our main result is then the following.

Theorem 1. *Let $\phi_\epsilon(x, t)$ denote the solution of the Cauchy initial-value problem for the MNLS equation (1) subject to initial data of the form (2) with $A_0(\cdot)$ and $S_0(\cdot)$ given by (11), where (without loss of generality) $\nu = 1$ and $S_0(0) = 0$, and where $\alpha > 0$, δ , and μ are subject to the inequalities (15)–(17). Then there exists a smooth curve $x = x_c(t)$, $t \geq 0$, with $x_c(0) = x_c$ such that for all $t \geq 0$ and all $x < x_c(t)$,*

$$(20) \quad \phi_\epsilon(x, t) = A(x, t)e^{iS(x, t)/\epsilon} + O(\epsilon), \quad \epsilon \rightarrow 0, \quad \epsilon > 0,$$

where the error term is uniform for (x, t) in compact subsets and where $A(x, t)$ and $S(x, t)$ are smooth, real-valued functions independent of ϵ that satisfy $A(x, 0) = A_0(x)$ and $S(x, 0) = S_0(x)$. Also, $Q < 0$ holds strictly for all $x < x_c(t)$ and $t \geq 0$, while $Q \rightarrow 0$ as $x \rightarrow x_c(t)$, $x < x_c(t)$. Finally, whenever $x < x_c(t)$ and $t > 0$, the Madelung-type fields

$$(21) \quad \rho(x, t) := A(x, t)^2 \quad \text{and} \quad u(x, t) := \frac{\partial S}{\partial x}(x, t)$$

exactly satisfy the dispersionless MNLS system (9).

The functions $A(x, t)$ and $S(x, t)$ will be specified precisely as part of the proof; see (220). Significantly, these functions are obtained without directly analyzing the dispersionless MNLS system, but rather arise from the solution of certain algebraic equations (the expressions given in (220) depend on the solution of (79) and the identities (155) and (160)). Our result is interesting in part because it is a *global* description of the solution $\phi_\epsilon(x, t)$ valid for all time $t \geq 0$ in a semi-infinite spatial domain $x < x_c(t)$. Unlike the semiclassical limit for the focusing nonlinear Schrödinger equation (for which the dispersionless system analogous to (9) is elliptic for all initial data) [14, 21], the solution of the MNLS equation with the initial data under consideration does not “break” at some caustic curve beyond which the Madelung ansatz fails and a higher-genus multiphase behavior ensues. Indeed, the boundary curve $x = x_c(t)$ is merely a *sonic line* across which the asymptotic solution changes from subsonic type (for $x < x_c(t)$) to supersonic type (for $x > x_c(t)$). While we do not give the proof in this paper, the sonic transition turns out not to be catastrophic for the Madelung fields, which extend smoothly into the supersonic region (although in this region the Madelung theory turns out to no longer be globally valid in time, that is, the MNLS solution “breaks” in the hyperbolic region for the dispersionless MNLS system). The analysis corresponding to the supersonic region $x > x_c(t)$ is the subject of our forthcoming paper [9] in this series.

From the point of view of technique, this paper contains several novelties that we wish to point out. The basic approach is to formulate the Riemann-Hilbert problem of inverse-scattering for the MNLS Cauchy problem with available explicit formulae for the scattering data (the reflection coefficient is the key quantity). Being as the formula for the reflection coefficient involves the Euler gamma function with arguments proportional to the large parameter ϵ^{-1} , one can hardly resist applying Stirling-type approximations. While much of the analysis corresponds to working with these approximations, we are very careful about not discarding any error terms until they can be rigorously controlled (which always occurs at the very end of the process in the steepest descent method). In particular, note that the function $\phi_\epsilon(x, t)$ in Theorem 1 *is the exact solution of the Cauchy initial-value problem for the specified initial data*; we do not replace the initial condition with a nearby function corresponding to approximate scattering data as is a common approach in the subject (see, for example [2], [14], and [21]). There are points where the Stirling approximation fails (these correspond to the exceptional values of the spectral parameter where turning points collide in WKB theory), and we deal with these points with the help of a new kind of “lens-opening” deformation of the Riemann-Hilbert problem that completely sidesteps the need for any local analysis near these points. This same technique also completely solves certain problems arising from the fact that in cases where one is lucky enough that the reflection coefficient admits an analytic continuation away from the real axis, there frequently exist “phantom poles” that do not correspond to eigenvalues of the spectral problem but nonetheless that obstruct the sort of contour deformations that are at the heart of the steepest descent method of Deift and Zhou.

In §2 we summarize the key facts about the scattering data for the WKI spectral problem in the case that the initial conditions for the MNLS equation are given in the special form (11), incidentally correcting a sign error in the original reference [8]. Then in §3 we recall the formulation of the inverse scattering problem for the MNLS equation as a Riemann-Hilbert problem of analytic matrix factorization. To prepare the Riemann-Hilbert problem for analysis in the semiclassical limit we need to introduce an appropriate “ g -function” $g(z)$ and a closely related function $h(z)$; these functions are constructed and completely characterized in §4. We should point out that it is the count of the zeros of the auxiliary function $Y(z)$ related to derivatives of g and h (see Proposition 1 in §4) that ultimately gives rise to the global-in-time nature of our main result. Finally, in §5 we use the g -function along with the new “lens-opening” method to reduce the Riemann-Hilbert problem formulated in §3 to a form where it can be treated by means of Neumann series for an associated

system of singular integral equations of small-norm type. The estimates resulting from this analysis complete the proof of Theorem 1.

Regarding notation, we use $\Re\{z\}$ and $\Im\{z\}$ to denote the real and imaginary parts, respectively, of a complex number z . We will typically use the symbol η to refer to an arbitrarily small positive quantity. We write all matrices in boldface (e.g. \mathbf{M}), with the exception of the Pauli spin matrices:

$$(22) \quad \sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

In particular, the expression $a^{b\sigma_3}$ appears frequently; this is simply the diagonal matrix

$$(23) \quad a^{b\sigma_3} := \begin{bmatrix} a^b & 0 \\ 0 & a^{-b} \end{bmatrix}, \quad a, b \in \mathbb{C}.$$

If $f(z)$ denotes a harmonic or analytic scalar or matrix-valued function defined on the complement of an oriented arc \mathcal{A} in some neighborhood of the complex z -plane, then for $z \in \mathcal{A}$ we define $f_+(z)$ (respectively $f_-(z)$) as the boundary value taken by $f(w)$ as $w \rightarrow z \in \mathcal{A}$ from the left (respectively right) as \mathcal{A} is traversed according to its orientation. Throughout our paper, when we write z^p for generally complex z and p we *always* mean the principal branch defined by $z^p := e^{p \log(z)}$ with $-\pi < \Im\{\log(z)\} < \pi$. Finally, we denote complex conjugation with an asterisk: z^* .

2. THE SCATTERING DATA AND PROPERTIES THEREOF

2.1. Exact Formulae. According to the theory of the inverse-scattering transform as described in the Appendix of [8], the key *scattering coefficients* needed to formulate the inverse-scattering problem are denoted $S_{12}(k)$ (defined for $\Im\{k^2\} = 0$) and $S_{22}(k)$ (defined for $\Im\{k^2\} \geq 0$ and analytic in the interior of this region). For initial data of the special form (2) with $A(\cdot)$ and $S(\cdot)$ given by (11), the linear scattering problem was reduced to a Gauss hypergeometric equation in [8], which allowed the relevant scattering coefficients to be calculated explicitly in terms of the Euler gamma function [1]. In the case of $\nu = 1$ and $S_0 = 0$ the scattering coefficients found in this way are exactly²

$$(24) \quad S_{12}(k) := \frac{2^{-i\mu/\epsilon} i\epsilon}{2k} \cdot \frac{\Gamma(\frac{1}{2} - \frac{i\mu}{2\epsilon} - \frac{\Omega}{\epsilon}) \Gamma(\frac{1}{2} - \frac{i\mu}{2\epsilon} + \frac{\Omega}{\epsilon})}{\Gamma(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon}) \Gamma(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})}, \quad \Im\{k^2\} = 0,$$

and

$$(25) \quad S_{22}(k) := \frac{\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} + \frac{i\mu}{2\epsilon}) \Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} - \frac{i\mu}{2\epsilon})}{\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} - \frac{R}{2\epsilon}) \Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} + \frac{R}{2\epsilon})}, \quad \Im\{k^2\} \geq 0.$$

Here

$$(26) \quad \Omega = \Omega(k) := \frac{2}{i\alpha} \left(k^2 - \frac{1 - \alpha\delta}{4} \right)$$

and³

$$(27) \quad R = R(k) := e^{i\pi/4} (-i(16k^2 - \mu^2))^{1/2}.$$

The corresponding *reflection coefficient* is defined generally in terms of the scattering coefficients as follows:

$$(28) \quad r(k) := -\frac{S_{12}(k)}{S_{22}(k)}, \quad k^2 \in \mathbb{R}.$$

In §3 we will show that we only need to consider k to lie on the boundary of the first quadrant in the complex plane (and for some purposes to admit analytic continuation into the open first quadrant), so we introduce a new coordinate z as follows:

$$(29) \quad z = k^2 \quad \text{and} \quad k = e^{i\pi/4} (-iz)^{1/2}.$$

²Equation (24) corrects a sign error in [8]. In that paper, the sign error originates in equation (146) and subsequently propagates into (151), (156), and (158)–(160).

³The function $R(k)$ defined in (27) is actually the Schwarz reflection of the function with the same name used in [8].

Thus, the closed upper half z -plane⁴ corresponds in a one-to-one fashion to the closed first quadrant in the k -plane. In terms of z , the specific reflection coefficient under consideration is then

$$(30) \quad s(z) := r(e^{i\pi/4}(-iz)^{1/2}) = -\frac{2^{-i\mu/\epsilon} i\epsilon}{2e^{i\pi/4}(-iz)^{1/2}} \frac{\Gamma(\frac{1}{2} - \frac{i\mu}{2\epsilon} - \frac{\Omega}{\epsilon})}{\Gamma(\frac{1}{2} + \frac{i\mu}{2\epsilon} + \frac{\Omega}{\epsilon})} \frac{\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} - \frac{R}{2\epsilon})}{\Gamma(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})} \frac{\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} + \frac{R}{2\epsilon})}{\Gamma(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})}, \quad z \in \mathbb{R},$$

where Ω and R are written explicitly in terms of $z = k^2$.

A simple exact formula is available for $|s(z)|^2$ for $z \in \mathbb{R}$ as a result of three observations. Firstly, since Ω is imaginary for $z \in \mathbb{R}$, and since $\Gamma(w^*) = \Gamma(w)^*$, it follows that

$$(31) \quad \left| \frac{\Gamma(\frac{1}{2} - \frac{i\mu}{2\epsilon} - \frac{\Omega}{\epsilon})}{\Gamma(\frac{1}{2} + \frac{i\mu}{2\epsilon} + \frac{\Omega}{\epsilon})} \right| = 1, \quad z \in \mathbb{R}.$$

Secondly, the function

$$(32) \quad m(z) := \frac{\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} - \frac{R}{2\epsilon})\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} + \frac{R}{2\epsilon})}{\Gamma(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})\Gamma(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})}$$

is meromorphic in a horizontal strip containing the real z axis, and hence so is the function $M(z) := m(z)m(z^*)^*$. Of course, for $z \in \mathbb{R}$ we have $M(z) = |m(z)|^2$, and we may calculate $M(z)$ for all real z by obtaining a formula valid for z in some interval of the real axis and applying analytic continuation to assert that the same formula holds for all $z \in \mathbb{R}$. For $z < \mu^2/16$, R and Ω are both imaginary, so using $\Gamma(w^*) = \Gamma(w)^*$ we obtain

$$(33) \quad M(z) = \frac{[\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} - \frac{R}{2\epsilon})\Gamma(\frac{1}{2} - \frac{\Omega}{\epsilon} + \frac{R}{2\epsilon})][\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} + \frac{R}{2\epsilon})\Gamma(\frac{1}{2} - \frac{\Omega}{\epsilon} - \frac{R}{2\epsilon})]}{[\Gamma(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})\Gamma(\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})][\Gamma(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})\Gamma(\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})]}, \quad z < \frac{\mu^2}{16}.$$

Thirdly, we recall the two reflection identities for the gamma function [1]:

$$(34) \quad \Gamma(w)\Gamma(-w) = -\frac{\pi}{w \sin(\pi w)} \quad \text{and} \quad \Gamma(\frac{1}{2} + w)\Gamma(\frac{1}{2} - w) = \frac{\pi}{\cos(\pi w)}.$$

Using these in (33), we obtain

$$(35) \quad \begin{aligned} M(z) &= \frac{(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon}) \sin(\pi(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})) \sin(\pi(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon}))}{\cos(\pi(\frac{\Omega}{\epsilon} - \frac{R}{2\epsilon})) \cos(\pi(\frac{\Omega}{\epsilon} + \frac{R}{2\epsilon}))} \\ &= \frac{\mu^2 - (-iR)^2}{4\epsilon^2} \frac{\cosh^2(\frac{\pi\mu}{2\epsilon}) - \cosh^2(\frac{\pi R}{2i\epsilon})}{\cosh^2(\frac{\pi\Omega}{i\epsilon}) + \cosh^2(\frac{\pi R}{2i\epsilon}) - 1}. \end{aligned}$$

It is clear that the latter is an even function of R and hence defines the analytic continuation of $M(z)$ to the whole real z -axis. From these three facts it therefore follows that

$$(36) \quad |s(z)|^2 = \text{sgn}(z) \frac{\cosh^2(\frac{\pi\mu}{2\epsilon}) - \cosh^2(\frac{\pi R}{2i\epsilon})}{\cosh^2(\frac{\pi\Omega}{i\epsilon}) + \cosh^2(\frac{\pi R}{2i\epsilon}) - 1} = \frac{|\cosh^2(\frac{\pi\mu}{2\epsilon}) - \cosh^2(\frac{\pi R}{2i\epsilon})|}{\cosh^2(\frac{\pi\Omega}{i\epsilon}) + \cosh^2(\frac{\pi R}{2i\epsilon}) - 1}, \quad z \in \mathbb{R}.$$

Since $\cosh^2(\frac{\pi R}{2i\epsilon})$ is even in R , it follows that the product $\text{sgn}(z)|s(z)|^2$ has a meromorphic continuation from \mathbb{R} to the whole complex z -plane.

2.2. Analytic Properties of the Scattering Data. When the scattering coefficients are given by the specific formulae (24)–(25), it is obvious that S_{12} and S_{22} are even functions of R and hence they have no branch point at $z = \mu^2/16$, the square-root branch point of R considered as a function of $z = k^2$. This makes $kS_{12}(k)$ and $S_{22}(k)$ both meromorphic functions of $z = k^2$. From this it follows that $s(z)$ has a meromorphic continuation from the real z -axis into the full upper half z -plane. The point $z = 0$ is a branch point of $s(z)$, however, as is evidenced by the fact that the product $s(z)(-iz)^{-1/2}$ is analytic and non-vanishing near $z = 0$ (the apparent singularity at $z = 0$ from the factor $(-iz)^{-1/2}$ is cancelled by the simple zero of $\Gamma(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})^{-1}$). This shows that $s(z)$ does not have any single-valued continuation into the lower half z -plane due to the mismatch of boundary values taken on the branch cut of $(-iz)^{1/2}$ along the negative imaginary axis. In this paper, we will refer to both the meromorphic continuation of $s(z)$ into the upper half z -plane as well as its

⁴The coordinate z defined in (29) is the negative of the variable z used in [8].

boundary value taken on the real z -axis by the same notation: $s(z)$. Now we consider the analytic nature of $s(z)$ for $\Im\{z\} \geq 0$.

In general, any zeros of the scattering coefficient $S_{22}(k)$ for $\Im\{k^2\} \geq 0$ (equivalently for $\Im\{z\} \geq 0$) represent discrete spectrum (eigenvalues) of the direct scattering problem and ultimately poles in the matrix unknown of the Riemann-Hilbert problem of inverse-scattering theory. For technical reasons we would like to avoid having to include such poles. This motivates the condition (16) on the parameters α , δ , and μ ; indeed in the case that $S_{22}(k)$ is given by the formula (25), it was shown in [8] that under the condition (16) there are no zeros in the closed upper half z -plane for any $\epsilon > 0$. In this case, S_{22} is an analytic function of z for $\Im\{z\} > 0$ that extends continuously to the real z -axis and that for each $\epsilon > 0$ is bounded away from zero.

The scattering coefficient S_{12} generally has no analytic continuation from the axes $\Im\{k^2\} = 0$ (or the real axis $\Im\{z\} = 0$). However, in the case of the special formula (24), it is clear that S_{12} can be continued into the upper half z -plane as a meromorphic function. The factors $\Gamma(\frac{1}{2} - \frac{i\mu}{2\epsilon} + \frac{\Omega}{\epsilon})$ and $\Gamma(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})$ are analytic and nonvanishing for $\Im\{z\} \geq 0$, but the factor $\Gamma(\frac{1}{2} - \frac{i\mu}{2\epsilon} - \frac{\Omega}{\epsilon})$ contributes to S_{12} an infinite array of simple poles with small spacing $\alpha\epsilon/2$ along the vertical ray $\Re\{z\} = z_P$ and $\Im\{z\} > 0$ where

$$(37) \quad z_P := \frac{1}{4}(1 - \alpha\delta + \alpha\mu).$$

These poles are singularities of $s(z)$ that unlike zeros of S_{22} do not correspond to eigenvalues of the direct scattering problem; in [8] they are called *phantom poles*. The presence of such poles in the reflection coefficient of an analogous inverse-scattering problem for the focusing nonlinear Schrödinger equation has been previously dealt with [21] by means of the installation of an implicit local parametrix; by contrast the deformations of the Riemann-Hilbert problem that we will carry out in §5 are specially designed to render this unfortunate feature of the reflection coefficient completely harmless without the use of any local parametrix at all. The remaining factor $\Gamma(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})$ can only contribute zeros to S_{12} . For technical reasons we wish to prevent these zeros from lying in the upper half z -plane, which is the purpose of the condition (17) on μ . Were this condition not satisfied, the factor in question would contribute to S_{12} infinitely many simple zeros lying along (and densely filling out as $\epsilon \downarrow 0$) a parabolic curve in the upper half-plane, and their presence would cause difficulties with the use of Stirling's formula for asymptotic analysis (see §2.3).

Under assumptions (16) and (17), $s(z)$ is analytic and non-vanishing in the open upper half-plane with the exception of the “phantom poles” along the line $\Re\{z\} = z_P$. Note also that the condition (16) implies that the zero $z = \mu^2/16$ of R is greater than the zero $z = (1 - \alpha\delta)/4$ of Ω , and that conditions (15) and (17) taken together imply that $z_P > 0$.

2.3. Asymptotic Properties of the Scattering Data.

2.3.1. *Asymptotic behavior of $|s(z)|^2$.* The three conditions (15)–(17) also determine the qualitative asymptotic behavior of $|s(z)|^2$ given by (36) for $z \in \mathbb{R}$ in the limit $\epsilon \downarrow 0$ as we will now show. We have $\Omega \in i\mathbb{R}$ for all $z \in \mathbb{R}$, and μ is a real constant. The relative size of the positive real (quadratic and constant, respectively) quantities

$$(38) \quad q(z) := (-i\Omega)^2 = \frac{1}{4\alpha^2}(4z - (1 - \alpha\delta))^2 \quad \text{and} \quad c(z) := \left(\frac{\mu}{2}\right)^2 = \frac{\mu^2}{4}$$

will therefore play a role in determining the asymptotic behavior of $|s(z)|^2$ for real z in the limit $\epsilon \downarrow 0$. If $z > \mu^2/16$, then $R \in \mathbb{R}$ and these two functions determine the asymptotic behavior completely; on the other hand, if $z < \mu^2/16$, then $R \in i\mathbb{R}$, and we also need to take into account the positive real (linear) quantity

$$(39) \quad l(z) := \left(-i\frac{R}{2}\right)^2 = \frac{\mu^2}{4} - 4z, \quad z < \frac{\mu^2}{16}.$$

Extending the definition of $l(z)$ by setting $l(z) := 0$ for $z \geq \mu^2/16$, we observe that the asymptotic behavior of $|s(z)|^2$ is determined by the maximum of $q(z)$, $l(z)$, and $c(z)$:

- For those z for which $c(z) = \max\{q(z), l(z), c(z)\}$, $|s(z)|^2$ is exponentially large in the limit $\epsilon \rightarrow 0$ (exponential growth of $s(z)$).
- For those z for which $l(z) = \max\{q(z), l(z), c(z)\}$, $|s(z)|^2 - 1$ is exponentially small in the limit $\epsilon \rightarrow 0$ (pure oscillation of $s(z)$).

- For those z for which $q(z) = \max\{q(z), l(z), c(z)\}$, $|s(z)|^2$ is exponentially small in the limit $\epsilon \rightarrow 0$ (exponential decay of $s(z)$).

It is easy to see that for $z < 0$, $l(z) > c(z)$, and that for $z > 0$, $l(z) < c(z)$, while $l(0) = c(0)$. It follows that there are only three possible scenarios for how the real z -axis may be partitioned into intervals of growth, pure oscillation, and decay of $s(z)$, as illustrated in Figure 2. Noting that the points $z = (1 - \alpha\delta \pm \alpha|\mu|)/4$

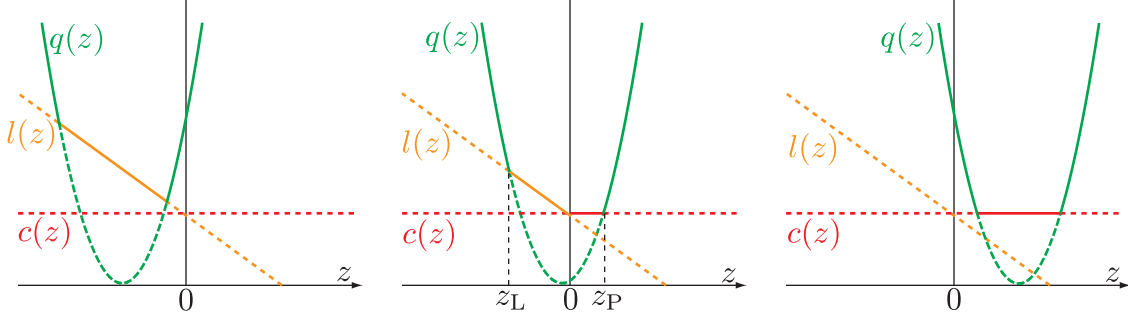


FIGURE 2. The three configurations of the reflection coefficient. Left: $1 - \alpha\delta + \alpha|\mu| < 0$. Center: $\alpha|\mu| > |1 - \alpha\delta|$. Right: $1 - \alpha\delta - \alpha|\mu| > 0$. The subdominant arcs of the graphs of $q(z)$, $l(z)$, and $c(z)$ are shown with dashed curves.

are the solutions of the equation $q(z) = c(z)$, these scenarios are:

- If $1 - \alpha\delta + \alpha|\mu| < 0$, then there are two real and negative roots of $q(z) = l(z)$, and for z between these two roots $s(z)$ is purely oscillatory, while for z outside of this interval $s(z)$ is exponentially small. See Figure 2, left-hand panel. In such cases one expects the asymptotic analysis of the Riemann-Hilbert problem of inverse scattering to resemble that which has been carried out for the Korteweg-de Vries equation in the small dispersion limit (see [6] as well as [3]–[5]). In such problems all of the necessary deformations of the Riemann-Hilbert problem that are required for the steepest descent technique are purely local to the real axis in the z -plane, and hence analyticity of the scattering data does not play any central role in the theory. This condition really corresponds to the case of globally supersonic initial data for the MNLS equation, for which there is a clear analogy with the inverse-scattering theory of the defocusing nonlinear Schrödinger equation in the semiclassical limit as shown in [10].
- If $1 - \alpha\delta - \alpha|\mu| > 0$, then $s(z)$ is exponentially large for $(1 - \alpha\delta - \alpha|\mu|)/4 < z < (1 - \alpha\delta + \alpha|\mu|)/4$ (where $c(z)$ is dominant) and otherwise is exponentially small. See Figure 2, right-hand panel. In such cases one expects that to handle the exponentially large reflection coefficient it will be necessary to introduce the meromorphic continuation of $s(z)$ from its interval of growth into the complex z -plane. This type of analysis was carried out for the semiclassical limit of the focusing nonlinear Schrödinger equation in [21]. In fact, this condition corresponds to the case of globally subsonic initial data for the MNLS equation, and the analogy with the focusing nonlinear Schrödinger equation in the semiclassical limit is also considered in [10].
- If $\alpha|\mu| > |1 - \alpha\delta|$, (that is, if (15) holds) then there is a unique negative root of $q(z) = l(z)$, given by $z = z_L := z^-$, where

$$(40) \quad z^\pm := \frac{1}{2} \left[\frac{1}{2}(1 - \alpha\delta) - \alpha^2 \pm \sqrt{\alpha^4 + \frac{\alpha^2 \mu^2}{4} - \alpha^2(1 - \alpha\delta)} \right],$$

and $s(z)$ is exponentially small for $z < z_L$ and for $z > (1 - \alpha\delta + \alpha|\mu|)/4$, is purely oscillatory for $z_L < z < 0$, and is exponentially large for $0 < z < (1 - \alpha\delta + \alpha|\mu|)/4$. See Figure 2, center panel. Note that under the assumption (17), the point of transition from exponential growth to exponential decay is exactly $z = z_P$.

These results show that the assumption (15), which was originally imposed to ensure that the initial conditions were of transsonic type, is also exactly what is required to ensure that in the spectral transform domain

there exist both a negative interval of pure oscillation and also an abutting positive interval of exponential growth of the reflection coefficient $s(z)$.

To handle the Riemann-Hilbert problem of inverse scattering in the transsonic case under consideration will therefore require in the same problem a combination of techniques from both the category of “modulationally stable” semiclassical limits (e.g. Korteweg-de Vries, defocusing nonlinear Schrödinger) and also the category of “modulationally unstable” semiclassical limits (e.g. focusing nonlinear Schrödinger).

Assuming the three conditions (15)–(17), note that as shown in the diagram in the central panel of Figure 2, there are five distinguished points on the real z -axis, in order from left-to-right:

- $z = z_L$, the point of transition from exponential decay to oscillatory behavior of $s(z)$. This is the negative root of the quadratic equation $q(z) = l(z)$.
- $z = z_\infty := \frac{1}{4}(1 - \alpha\delta - \alpha\mu)$, a point at which two subdominant exponentials in $|s(z)|^2$ exchange roles. This point is the negative root of the quadratic equation $q(z) = c(z)$ and it also has significance with respect to the *turning point curve* to be explained in §4.1.
- $z = 0$, the point of transition from oscillatory behavior to exponential growth of $s(z)$. This is the unique root of the linear equation $l(z) = c(z)$.
- $z = z^+$, another point at which two subdominant exponentials in $|s(z)|^2$ exchange roles. This point is the positive root of the quadratic equation $q(z) = l(z)$ and is given explicitly by (40).
- $z = z_P := \frac{1}{4}(1 - \alpha\delta + \alpha\mu)$, the point of transition from exponential growth to exponential decay of $s(z)$. This is the positive root of the quadratic equation $q(z) = c(z)$.

That z^\pm are real and distinct follows from condition (16) along with $\alpha > 0$, and that they have opposite signs is exactly equivalent to the condition (15). Note also that conditions (15) and (17) together imply that $z^+ < z_P$. For later convenience, let us define the linear exponent $f(z)$ by

$$(41) \quad f(z) := \frac{2\pi}{\alpha}(z - z_P).$$

The assumptions (15)–(17) imply that simple bounds for $|s(z)|^2$ then follow from the exact formula (36) and the central graph in Figure 2. Indeed, since $\cosh^2(\pi R/2i\epsilon) \geq \cosh^2(\pi\mu/2\epsilon)$ and $\cosh^2(\pi\Omega/i\epsilon) - 1 \geq 0$ for $z \leq 0$, we easily obtain the inequality:

$$(42) \quad |s(z)|^2 \leq 1, \quad z \leq 0.$$

On the other hand, for $z \geq 0$ we have instead that

$$(43) \quad \begin{aligned} |s(z)|^2 &\leq \frac{\cosh^2(\frac{\pi\mu}{2\epsilon})}{\cosh^2(\frac{\pi\Omega}{i\epsilon})} \\ &= e^{4\pi(\alpha\mu/4 - |z - (1 - \alpha\delta)/4|)/(\alpha\epsilon)} \left(\frac{1 + e^{-\pi\mu/\epsilon}}{1 + e^{-4\pi|z - (1 - \alpha\delta)/4|/(\alpha\epsilon)}} \right)^2 \\ &\leq 4e^{4\pi(\alpha\mu/4 - |z - (1 - \alpha\delta)/4|)/(\alpha\epsilon)} \\ &\leq 4e^{4\pi(z_P - z)/(\alpha\epsilon)} \\ &= 4e^{-2f(z)/\epsilon}, \quad z \geq 0. \end{aligned}$$

Note that by the Mean Value Theorem applied to (36), the estimates (42) and (43) can be replaced by an improved estimate valid in a neighborhood of the origin as follows: for each sufficiently small $\eta > 0$ there exist constants $C > 0$ and $K > 0$ independent of ϵ such that

$$(44) \quad |s(z)|^2 \leq \frac{C|z|}{\epsilon} e^{-K/\epsilon}, \quad |z| < \eta.$$

Here we have used the fact that $z_P > 0$ to ensure that $K > 0$.

Similar reasoning produces the following asymptotic formulae involving $|s(z)|^2$. Firstly, we have

$$(45) \quad 1 - |s(z)|^2 = e^{-\tau(z)/\epsilon} (1 + \text{exponentially small in } \epsilon), \quad \text{uniformly for } z \in [z_L + \eta, z_\infty - \eta] \cup [z_\infty + \eta, 0]$$

for all $\eta > 0$, where $\tau(z) > 0$ is defined by

$$(46) \quad \tau(z) = \begin{cases} \tau_{z_L}(z) := -i\pi R - 2\pi|\Omega| = \pi\sqrt{\mu^2 - 16z} + \frac{4\pi}{\alpha} \left(z - \frac{1 - \alpha\delta}{4} \right), & z_L < z \leq z_\infty \\ \tau_0(z) := -i\pi R - \pi\mu = \pi\sqrt{\mu^2 - 16z} - \pi\mu, & z_\infty \leq z < 0. \end{cases}$$

The function $\tau(z)$ extends (by zero) to a continuous function for $z \in \mathbb{R}$. Secondly, we have

$$(47) \quad |s(z)|^2 = e^{-2f(z)/\epsilon} (1 + \text{exponentially small in } \epsilon), \quad \text{uniformly for } z > z^+ + \eta, \forall \eta > 0.$$

2.3.2. Stirling asymptotics for $s(z)$. Under conditions (15)–(17), Stirling's formula [1] yields accurate asymptotics for $s(z)$ that are uniformly accurate to a relative error of order $O(\epsilon)$ in the closed set S consisting of the closed upper half-plane with the following sets omitted:

- The vertical strip $|\Re\{z\} - z_P| < \eta$ with $\Im\{z\} \geq 0$. This strip contains the accumulation locus of the phantom poles. The Stirling approximation of the ratio $\Gamma(\frac{1}{2} - \frac{i\mu}{2\epsilon} - \frac{\Omega}{\epsilon})/\Gamma(\frac{1}{2} + \frac{i\mu}{2\epsilon} + \frac{\Omega}{\epsilon})$ fails (only) in this strip as $\epsilon \downarrow 0$.
- The open upper half-disk of radius η centered at $z = 0$. The point $z = 0$ is the unique root of $R - i\mu$ for $\Im\{z\} \geq 0$ when $\mu > 0$ (and $R + i\mu$ is nonvanishing for $\Im\{z\} \geq 0$). The Stirling approximation of the product $\Gamma(-\frac{i\mu}{2\epsilon} - \frac{R}{2\epsilon})\Gamma(-\frac{i\mu}{2\epsilon} + \frac{R}{2\epsilon})$ fails (only) in this half-disk as $\epsilon \downarrow 0$.
- The open upper half-disks of radius η centered at the real points $z^- < 0 < z^+$ defined by (40). These are the roots of the quadratic equation $R^2 = (2\Omega)^2$. The Stirling approximation of the product $\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} - \frac{R}{2\epsilon})\Gamma(\frac{1}{2} + \frac{\Omega}{\epsilon} + \frac{R}{2\epsilon})$ fails (only) in the union of these half-disks as $\epsilon \downarrow 0$.

Here η is any arbitrarily small fixed value. The omitted vertical strip containing the phantom poles divides S into two disjoint subregions: S_L to the left of the strip and S_R to the right. In each of these two subregions Stirling's formula will supply a different analytic approximation to the ratio $s(z)$. The resulting formulae are as follows. Firstly, define (here $\log(\cdot)$ denotes the principal branch with $|\Im\{\log(\cdot)\}| < \pi$)

$$(48) \quad \begin{aligned} \Phi(z) := & -\mu \log(2) + \frac{4}{\alpha} (z - z_P) \log\left(\frac{2}{\alpha}(z_P - z)\right) \\ & + \frac{\mu - iR}{2} \log\left(\frac{-i\mu - R}{2}\right) + \frac{\mu + iR}{2} \log\left(\frac{-i\mu + R}{2}\right) \\ & - i\left(\Omega - \frac{R}{2}\right) \log\left(\Omega - \frac{R}{2}\right) - i\left(\Omega + \frac{R}{2}\right) \log\left(\Omega + \frac{R}{2}\right), \quad \Im\{z\} > 0. \end{aligned}$$

Under the conditions (15)–(17), $\Phi(z)$ is an analytic function of z for $\Im\{z\} > 0$. Now, define:

$$(49) \quad E(z) := -s(z)e^{-i\Phi(z)/\epsilon}, \quad \text{for } \Im\{z\} > 0.$$

The function $E(z)$ is meromorphic for $\Im\{z\} > 0$ and its poles are confined to the line $\Re\{z\} = z_P$. In particular it is analytic in the interior of $S = S_L \cup S_R$. Stirling's formula implies that

$$(50) \quad E(z) = \begin{cases} 1 + O(\epsilon), & z \in S_L, \\ e^{-2f(z)/\epsilon} (1 + O(\epsilon)), & z \in S_R, \end{cases}$$

where $f(z)$ is defined by (41), with both estimates holding uniformly in the indicated region (including for the boundary value $E_+(z)$ taken on the real axis). Next, define

$$(51) \quad \tilde{E}(z) := \left(1 + e^{2f(z)/\epsilon}\right) E(z), \quad \Im\{z\} > 0.$$

The explicit prefactor exactly cancels the poles in $E(z)$ for $\Im\{z\} > 0$, and hence $\tilde{E}(z)$ is an analytic function in the whole upper half complex plane. Letting \tilde{S} denote the closed upper half-plane with the open upper half-disks of radius η centered at $z = z_L$, $z = 0$, $z = z^+$, and $z = z_P$ omitted, we can obtain from Stirling's formula (by using first the reflection identities (34) for $\Gamma(\cdot)$) that

$$(52) \quad \tilde{E}(z) = 1 + O(\epsilon), \quad z \in \tilde{S}.$$

Again, the estimate holds uniformly (including for the boundary value $\tilde{E}_+(z)$). It will be important later to record the imaginary part of the boundary values taken by Φ on the real axis:

$$(53) \quad \Im\{\Phi_+(z)\} = \begin{cases} \frac{i\pi}{2} (R(z) - 2\Omega(z)), & z < z_L = z^- \\ 0, & z_L = z^- < z < 0 \\ -\frac{i\pi}{2} (R(z) - i\mu), & 0 < z < z^+ \\ -\frac{2\pi}{\alpha} |z - z_P| = -|f(z)|, & z > z^+. \end{cases}$$

It therefore follows that $\Phi(z)$ extends through the interval $z_L < z < 0$ to an analytic function defined also in the open lower half-plane, and henceforth we consider $\Phi(z)$ to be analytic in the slit domain $z \in \mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty))$ and to satisfy the Schwarz symmetry condition $\Phi(z^*) = \Phi(z)^*$. By direct calculation, one sees from (48) that

$$(54) \quad \Phi'(z) = -\frac{2\pi i}{\alpha} - \pi z^{-1/2} - \frac{\mu}{2} z^{-1} + O(z^{-3/2}), \quad z \rightarrow \infty, \quad \Im\{z\} > 0.$$

An important observation is that while here the formulae (46) and (48) for $\tau(z)$ and $\Phi(z)$ respectively come directly from analysis of the exact formula (30) for the reflection coefficient valid in the case of the particular initial conditions under consideration, these formulae agree *exactly* with predictions based on WKB theory formally valid for more general initial data. This formal semiclassical spectral analysis of the direct scattering problem can be found in all details in §2 of [10]. The more general formulae corresponding to (46) and (48) arise from WKB theory as follows. The *turning point curve* of WKB theory is, by definition, the locus of real x and complex z satisfying the characteristic equation

$$(55) \quad \chi(x; z) := (4z - 1 + \alpha u_0(x))^2 + 16\alpha^2 z \rho_0(x) = 0.$$

Let $z \in \mathbb{R}$ be a real value for which the function $\chi(x; z)$ has real roots x , and let $x_-(z)$ denote the most negative of these roots. In particular, $\chi(x; z) > 0$ for all $x < x_-(z)$. Define signs σ_\pm by

$$(56) \quad \sigma_\pm := \operatorname{sgn}(4z - 1 + \alpha u_\pm),$$

and then set

$$(57) \quad \omega(x; z) := \sigma_- \sqrt{\chi(x; z)}, \quad x < x_-(z).$$

For each appropriate value of $z \in \mathbb{R}$, this quantity has the fixed sign σ_- throughout its domain of definition, and $\omega(x; z) \rightarrow \omega_-(z)$ as $x \rightarrow -\infty$, where

$$(58) \quad \omega_-(z) := 4z - 1 + \alpha u_-.$$

Then, for all real z for which $\chi(x; z)$ has real roots, we let

$$(59) \quad \gamma(x; z) := \sqrt{-\chi(x; z)} > 0, \quad \chi(x; z) \leq 0,$$

and then

$$(60) \quad \tau(z) = \frac{1}{\alpha} \int_{y: \chi(y; z) < 0} \gamma(y; z) dy$$

is a generalized formula that agrees exactly with (46) when $A_0(x)$ and $S_0(x)$ are the particular functions (11) considered in this paper. Next, we have

$$(61) \quad \Phi(z) = \frac{1}{\alpha} \int_{-\infty}^{x_-(z)} [\omega(y; z) - \omega_-(z)] dy + \frac{1}{\alpha} \omega_-(z) x_-(z) + S_-.$$

This formula is a generalization of (48) in the sense that the latter is the analytic continuation to the upper half z -plane of (61) in the special case of the functions $A_0(x)$ and $S_0(x)$ given by (11). For these functions it turns out that there are generically either two distinct roots (turning points) $x_-(z) < x_+(z)$ of $\chi(x; z)$ or none for $z \in \mathbb{R}$. In such cases it can be shown that if z_0 is a real value at which $x_\pm(z)$ coalesce in a

square-root sense, then the functions $\Phi(z)$ and $\tau(z)$ can be analytically continued about such a point, and the following monodromy relations hold:

$$(62) \quad \Phi(z_0 + (z - z_0)e^{\pm 2\pi i}) = \Phi(z) \pm i\sigma_- \tau(z),$$

$$(63) \quad \tau(z_0 + (z - z_0)e^{\pm 2\pi i}) = \tau(z)$$

We will make use of the concrete formulae (46) and (48) and also the equivalent WKB formulae (60) and (61) at various points in our analysis.

3. RIEMANN-HILBERT PROBLEM OF INVERSE SCATTERING

3.1. Basic problem. We take as a starting point the formulation of the inverse-scattering problem as a matrix Riemann-Hilbert problem for an unknown $\mathbf{M}(k; x, t)$ as described in the Appendix of [8]. This matrix has jump discontinuities across the real and imaginary axes in the complex k -plane and satisfies the holomorphic involution $\mathbf{M}(-k; x, t) = i^{\sigma_3} \mathbf{M}(k; x, t) i^{-\sigma_3}$ and the antiholomorphic involution $\mathbf{M}(-k^*; x, t)^* = \sigma_1 \mathbf{M}(k; x, t) \sigma_1$. If for $\Im\{k^2\} = 0$, $\mathbf{M}_{\pm}(k; x, t)$ denotes the boundary value taken by \mathbf{M} from a quadrant where $\pm\Im\{k^2\} < 0$, then

$$(64) \quad \mathbf{M}_{+}(k; x, t) = \mathbf{M}_{-}(k; x, t) e^{i\theta(k^2; x, t)\sigma_3/\epsilon} \begin{bmatrix} 1 \pm |r(k)|^2 & r(k) \\ \pm r(k)^* & 1 \end{bmatrix} e^{-i\theta(k^2; x, t)\sigma_3/\epsilon}, \quad \pm k^2 > 0,$$

where

$$(65) \quad \theta(z) = \theta(z; x, t) := -\frac{2}{\alpha} \left(z - \frac{1}{4} \right) x - \frac{4}{\alpha^2} \left(z - \frac{1}{4} \right)^2 t$$

and where $r(k) = -r(-k)$ is the reflection coefficient defined generally in terms of the scattering coefficients $S_{12}(k)$ and $S_{22}(k)$ by (28). The matrix $\mathbf{M}(k; x, t)$ is normalized to the identity at $k = 0$:

$$(66) \quad \lim_{k \rightarrow 0} \mathbf{M}(k; x, t) = \mathbb{I}$$

regardless of the quadrant from which the limit is taken, and is required to have a well-defined limiting value (a matrix-valued function of (x, t)) as $k \rightarrow \infty$. In the case under study in this paper that there are no eigenvalues of the scattering problem, the matrix $\mathbf{M}(k; x, t)$ is analytic for $\Im\{k^2\} \neq 0$ and the above properties uniquely determine $\mathbf{M}(k; x, t)$ given $r(\cdot)$. The corresponding solution of the MNLS equation is then obtained from $\mathbf{M}(k; x, t)$ by the formula

$$(67) \quad \phi_{\epsilon}(x, t) := \lim_{k \rightarrow \infty} \frac{2k}{\alpha} \frac{M_{12}(k; x, t)}{M_{22}(k; x, t)}.$$

(Note that $\mathbf{M}(k; x, t)$ is asymptotically diagonal as $k \rightarrow \infty$. See [8, page 990].)

3.2. Symmetry reduction. It will be convenient to take advantage of the $k \mapsto -k$ symmetry to define a new unknown by setting $\mathbf{N}(z; x, t) := \mathbf{M}(i(-z)^{1/2}; x, t)$ for $\Im\{z\} \neq 0$. Since $(-z)^{1/2}$ denotes the principal branch of the square root, the matrix $\mathbf{M}(k; x, t)$ as defined in the first (second) quadrant of the k -plane is equivalent to the matrix $\mathbf{N}(z; x, t)$ in the upper (lower) half z -plane. For $z \in \mathbb{R}$, let $\mathbf{N}_{\pm}(z; x, t)$ denote the boundary value of \mathbf{N} taken from \mathbb{C}_{\pm} . Then, if $z > 0$, since $\mathbf{N}_{\pm}(z; x, t) = \mathbf{M}_{\mp}(\pm z^{1/2}; x, t)$ and according to the holomorphic symmetry of $\mathbf{M}(k; x, t)$ we have $\mathbf{M}_{+}(-z^{1/2}; x, t) = i^{\sigma_3} \mathbf{M}_{+}(z^{1/2}; x, t) i^{-\sigma_3}$,

$$(68) \quad \mathbf{N}_{+}(z; x, t) = i^{\sigma_3} \mathbf{N}_{-}(z; x, t) i^{-\sigma_3} e^{i\theta(z; x, t)\sigma_3/\epsilon} \begin{bmatrix} 1 & -r(z^{1/2}) \\ -r(z^{1/2})^* & 1 + |r(z^{1/2})|^2 \end{bmatrix} e^{-i\theta(z; x, t)\sigma_3/\epsilon}, \quad z > 0.$$

On the other hand, if $z < 0$, then $\mathbf{N}_{\pm}(z; x, t) = \mathbf{M}_{\mp}(i(-z)^{1/2}; x, t)$, so

$$(69) \quad \mathbf{N}_{+}(z; x, t) = \mathbf{N}_{-}(z; x, t) e^{i\theta(z; x, t)\sigma_3/\epsilon} \begin{bmatrix} 1 & -r(i(-z)^{1/2}) \\ r(i(-z)^{1/2})^* & 1 - |r(i(-z)^{1/2})|^2 \end{bmatrix} e^{-i\theta(z; x, t)\sigma_3/\epsilon}, \quad z < 0.$$

The function $e^{i\pi/4}(-iz)^{1/2}$ (principal branch) is analytic in the upper half z -plane and agrees with $z^{1/2}$ for $z > 0$ and with $i(-z)^{1/2}$ for $z < 0$. Note that both $r(z^{1/2})$ for $z > 0$ and $r(i(-z)^{1/2})$ for $z < 0$ correspond to the same function $s(z)$ defined for all $z \in \mathbb{R}$ by (30). We frequently omit the explicit dependence on

the parameters (x, t) (along with ϵ , α , δ , and μ) and simply write $\mathbf{N}(z) = \mathbf{N}(z; x, t)$. The matrix $\mathbf{N}(z)$ necessarily satisfies the Schwarz symmetry property:

$$(70) \quad \mathbf{N}(z^*) = \sigma_1 \mathbf{N}(z)^* \sigma_1$$

as well as the normalization condition

$$(71) \quad \lim_{\substack{z \rightarrow 0 \\ \Im\{z\} \neq 0}} \mathbf{N}(z) = \mathbb{I},$$

and we require $\mathbf{N}(z)$ to have a well-defined (necessarily diagonal) limiting value as $z \rightarrow \infty$.

4. THE FUNCTIONS $g(z)$ AND $h(z)$

In this section, we construct two related analytic functions of z that will depend parametrically on $(x, t) \in \mathbb{R}^2$. For notational convenience, we will frequently not write the parameters explicitly as arguments of these functions, but the reader should be aware that the dependence is there. We wish to stress from the outset that *all objects considered in this section are independent of the basic asymptotic parameter $\epsilon > 0$.*

4.1. The WKB turning point curve. In the case of the special initial data under consideration the turning point curve defined in general by (55) is given by the equation

$$(72) \quad (Z - A + BT)^2 + Z(1 - T^2) = 0, \quad z = \alpha^2 Z, \quad T := \tanh(x),$$

where the parameters A and B are defined by (18). This is quadratic in Z for fixed T and also is quadratic in T for fixed Z . In solving for Z , one easily finds that the discriminant is $-(1 - T^2)(T^2 - 4BT + 4A - 1)$, whose second factor is proportional by a negative constant to the quantity Q at $t = 0$ as defined by (14). Therefore under the conditions (15)–(17), the only (simple) root in the range $|T| < 1$ is $T = T_c$ defined by (19). (The other root of the quadratic factor exceeds $T = 1$ for (A, B) in the admissible region shown in Figure 1.) Moreover, the discriminant is positive (only) for $T \in (T_c, 1)$. The point $x_c = \operatorname{arctanh}(T_c)$, corresponding to $z = z_c := (\alpha(\delta + \mu T_c) - 1)/4 = (\alpha u_0(x_c) - 1)/4$, is the only point on the turning point curve in the (x, z) -plane where there is a vertical tangent. For $x \geq x_c$ the turning point curve has two real branches that we denote as $z = \mathfrak{a}(x)$ and $z = \mathfrak{b}(x)$ with $\mathfrak{a}(x) \leq \mathfrak{b}(x)$, with equality holding only at $x = x_c$ and in the limit $x \rightarrow +\infty$ as illustrated in Figure 3.

Horizontal tangents to the turning point curve correspond to simultaneous solutions of (72) and the identity

$$(73) \quad B(Z - A + BT) - ZT = 0$$

obtained from (72) by differentiating implicitly with respect to T and setting $dZ/dT = 0$. Using this latter relation to eliminate Z from (72) we obtain an equation for T -values of horizontal tangents in the form

$$(74) \quad \frac{BT - A}{(T - B)^2} [(B^2 - A)T^2 + BT - B^2] = 0$$

Therefore, either $T = A/B$ or $T = T_{\pm}(A, B)$ where

$$(75) \quad T_{\pm}(A, B) := \pm \frac{2B}{\sqrt{4B^2 - 4A + 1} \pm 1}.$$

Since $B \geq 0$ and $B^2 \geq A$ both hold throughout the domain of values of (A, B) consistent with conditions (15)–(17) (see Figure 1), it is clear that $T_+(A, B) \geq 0$ while $T_-(A, B) \leq 0$. Moreover, $T_+(A, B) = 1$ implies that $B = A$ while $T_-(A, B) = -1$ implies that $B = -A$, both conditions that only occur on the boundary of the admissible region for (A, B) . Also, from the asymptotic expansions of $T_{\pm}(A, B)$ in the limit $B \uparrow \infty$ with $A \in \mathbb{R}$ fixed one concludes that $0 \leq T_+(A, B) < 1$ and $T_-(A, B) < -1$ both hold in the interior of the admissible region, and therefore of $T_{\pm}(A, B)$ only $T_+(A, B)$ corresponds to a value of $x \in \mathbb{R}$. Since $B > |A|$ holds in the interior of the admissible region, $T = A/B$ also corresponds to a real value of x . Similar analysis proves that both $T = A/B$ and $T = T_+(A, B)$ exceed T_c for (A, B) in the admissible region, so only the real part of the turning point curve for $T > T_c$ has any horizontal tangents (defined by $dZ/dT = 0$), and there are exactly two of them. Also, it is easy to check that throughout the admissible region of the (A, B) -plane, the horizontal tangent at $T = A/B$ occurs on the more positive of the two branches of $Z(T)$ (with value

$z = 0$) while the horizontal tangent at $T = T_+(A, B)$ occurs on the more negative of the two branches (with value z_L). We define x_1 and x_2 by

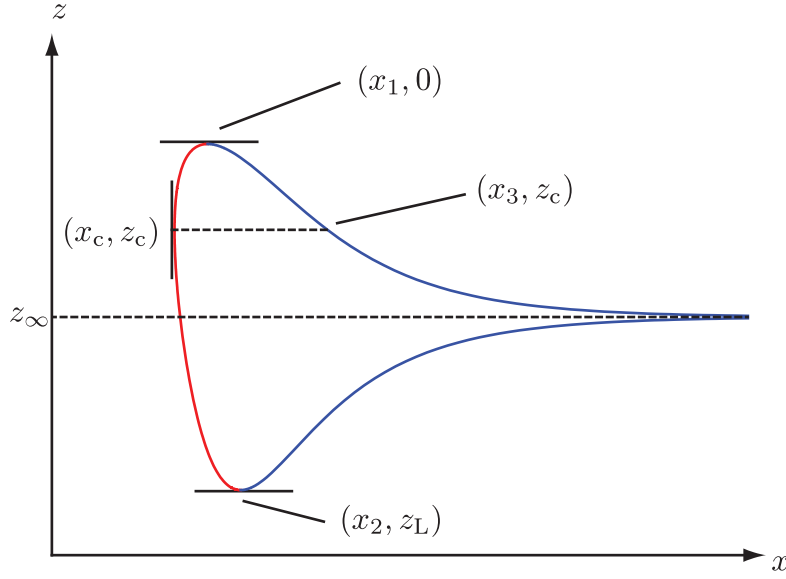


FIGURE 3. The real part of the turning point curve for (A, B) in the admissible region described by conditions (15)–(17). The red part of the curve is the graph of a function $x = x_-(z)$ and the blue part of the curve is the graph of a function $x = x_+(z)$ both over the interval $z_L < z < 0$. For such z , $x_{\pm}(z)$ are real turning points of WKB theory. The upper curve is the graph of a function $z = \mathfrak{b}(x)$ and the lower curve is the graph of a function $z = \mathfrak{a}(x)$, both defined for $x \geq x_c$. The singularity of $x_+(z)$ occurs when $Z = A - B$ or $z = z_{\infty} := (1 - \alpha\delta - \alpha\mu)/4$, the break point of the equation (46) for $\tau(z)$ in the interval $z_L < z < 0$.

$$(76) \quad x_1 := \operatorname{arctanh}\left(\frac{A}{B}\right) \quad \text{and} \quad x_2 := \operatorname{arctanh}(T_+(A, B)).$$

Note that under the conditions (15)–(17), we have the inequalities $x_2 > x_1 > x_c$. It is also possible to show that these conditions imply that $z_c \geq z_{\infty} := (1 - \alpha\delta - \alpha\mu)/4$ (z_{∞} as originally defined in §2.3 is also the point at which the branches of the turning point curve coalesce in the limit $x \rightarrow +\infty$, or equivalently, the singularity of $x_+(z)$), with equality occurring only along the part of the boundary of the admissible region of the parameter space illustrated in Figure 1 with $A \geq 0$. This in turn implies that there exists exactly one value of $x > x_1$ that we denote by x_3 , at which point $\mathfrak{b}(x_3) = z_c$. For $x_c < x < x_3$ we have $\mathfrak{a}(x) < z_c < \mathfrak{b}(x)$ and for $x > x_3$ we have $\mathfrak{a}(x) < \mathfrak{b}(x) < z_c$. It also follows from the conditions (15)–(17) that in fact $x_3 > x_2$. These relationships are illustrated in Figure 3.

For $x < x_c$, the turning point curve is complex, with two distinct complex-conjugate branches: $z = \mathfrak{z}(x)$ with $\Im\{\mathfrak{z}(x)\} > 0$ and $z = \mathfrak{z}(x)^*$. The two branches coincide (with a real value) only in the limits $x \uparrow x_c$ (with value $z = z_c$) and $x \downarrow -\infty$ (with value $Z = A + B > 0$ or $z = z_P$). Moreover, it is easy to prove that for (A, B) in the admissible region, $\Re\{\mathfrak{z}(x)\}$ is monotone decreasing in x for $-\infty < x < x_c$, and

$$(77) \quad \sup_{x < x_c} \Re\{\mathfrak{z}(x)\} = \lim_{x \rightarrow -\infty} \Re\{\mathfrak{z}(x)\} = z_P.$$

In particular, $\Re\{\mathfrak{z}(x)\} < z_P$ for all finite $x < x_c$. Note also that for $x < x_c$,

$$(78) \quad \Re\{\mathfrak{z}(x)\} = -\frac{1}{4} (\alpha u_0(x) - 1 + 2\alpha^2 \rho_0(x)).$$

4.2. Basic construction of $g(z)$ and $h(z)$. The auxiliary function $Y(z)$. Let a complex number q with $\Im\{q\} > 0$ be given, along with a simple oriented arc \mathcal{B} from $z = q$ to $z = z_P$ in the open upper half-plane. Let $S(z) = S(z; q, q^*, \mathcal{B})$ be the function uniquely defined by the following properties:

- $S(z)$ is defined and analytic for $z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*)$.
- $S(z)^2 = (z - q)(z - q^*)$.
- $S(z) = z + O(1)$ as $z \rightarrow \infty$.

At times we will want to think of q and q^* as being independent complex variables, but when they are linked by complex conjugation, $S(z)$ is obviously a Schwarz-symmetric function: $S(z^*) = S(z)^*$. Note that $S(z)$ changes sign across its branch cut $\mathcal{B} \cup \mathcal{B}^*$.

Let us assume (this will be completely clarified in §4.3) that given $(x, t) \in \mathbb{R}^2$, the complex number q satisfies the equations

$$(79) \quad \begin{aligned} M_0(q, q^*; x, t) &:= \frac{2\pi}{\alpha^2} (2(q + q^*)t + \alpha x - t) + I_0(q, q^*) = 0 \\ M_1(q, q^*; x, t) &:= \frac{\pi}{\alpha^2} ((3q^2 + 2|q|^2 + 3q^{*2})t + (q + q^*)(\alpha x - t)) + I_1(q, q^*) = 0, \end{aligned}$$

where

$$(80) \quad I_p(q, q^*) := \Im \left\{ \int_{\mathcal{B}} \frac{\Phi'(s)s^p ds}{S_+(s)} \right\}, \quad p = 0, 1.$$

As usual, the notation $S_+(z)$ denotes the boundary value taken from the left side of \mathcal{B} as the arc is traversed according to its orientation from q to z_P . Now, let $p(z)$ be the function analytic for $z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*)$ given by the Cauchy-type integral formula

$$(81) \quad p(z) := \frac{S(z)}{2\pi i} \left[\int_{\mathcal{B}} \frac{\Phi'(s) ds}{S_+(s)(s - z)} - (*)^* \right] + \theta'(z) + \frac{8tS(z)}{\alpha^2}, \quad z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*),$$

where the notation $(*)^*$ indicates the Schwarz reflection $w(z^*)^*$ of the function $w(z)$ immediately preceding the minus sign, and where $\theta(z) = \theta(z; x, t)$ is defined by (65). The conditions (79) obviously imply that

$$(82) \quad p(z) = O\left(\frac{1}{z^2}\right), \quad z \rightarrow \infty,$$

and hence the contour integral

$$(83) \quad g(z) := \int_0^z p(s) ds, \quad z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*)$$

is independent of path (as long as the path avoids the branch cut $\mathcal{B} \cup \mathcal{B}^*$ of p) and defines a function analytic in the same domain as p , with a well-defined limiting value $g(\infty)$. In terms of $g(z)$ we define a related function h as follows:

$$(84) \quad h(z) := \theta(z) + \frac{1}{2}\Phi(z) - g(z), \quad z \in \mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty) \cup \mathcal{B} \cup \mathcal{B}^*).$$

Because $\theta(z)$ is entire, $h(z)$ is analytic exactly in the intersection of the domains of analyticity of $\Phi(z)$ and $g(z)$ as explicitly indicated. Both g and h are Schwarz-symmetric functions: $g(z^*) = g(z)^*$ and $h(z^*) = h(z)^*$.

Let the function $Y(z) = Y(z; q, q^*)$ be defined for $\Im\{z\} > 0$ and z near \mathcal{B} by the following Cauchy-type integral:

$$(85) \quad Y(z; q, q^*) := \left[\frac{1}{2\pi i} \oint_L \frac{\Phi'(s) ds}{S(s)(s - z)} + (*)^* \right] + \frac{16t}{\alpha^2},$$

where L is a clockwise (negatively) oriented loop in the upper half s -plane beginning and ending at z_P and encircling both $s = z$ and the arc \mathcal{B} exactly once. Obviously Y is an analytic function of z in the interior of the loop L . Exploiting more detailed information about Φ (as is available from the specific formula (48) but that may or may not be available for more general initial data when one uses instead the WKB formula (61)) allows us define a global analytic continuation of $Y(z; q, q^*)$ as we will now show. Indeed, using the fact that $\Phi'(z)$ is analytic in the open upper half-plane and satisfies $\Phi'(z^*)^* = \Phi'(z)$, along with the asymptotic

estimate $\Phi'(z) = O(\log|z|)$ as $z \rightarrow \infty$ allows us to deform the contour L to a contour along the real axis; therefore

$$(86) \quad Y(z; q, q^*) = \frac{16t}{\alpha^2} - \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{\Phi'_+(s) - \Phi'_-(s)}{S(s)(s-z)} ds = \frac{16t}{\alpha^2} - \frac{1}{\pi} \int_{\mathbb{R}} \frac{\Im\{\Phi'_+(s)\}}{S(s)(s-z)} ds$$

defines the analytic continuation of Y to the domain $z \in \mathbb{C} \setminus \mathbb{R}$. Now, we recall (53) along with the definitions (26) of $\Omega(z)$ and (27) of $R(z)$ and the fact that for $s \in \mathbb{R}$, $S(s) = -\sqrt{(s - \Re\{q\})^2 + \Im\{q\}^2}$ for $s < z_P$ and $S(s) = \sqrt{(s - \Re\{q\})^2 + \Im\{q\}^2}$ for $s > z_P$, and we therefore find that when Φ is given by (48),

$$(87) \quad Y(z; q, q^*) := \frac{16t}{\alpha^2} - \frac{2}{\alpha} \int_{-\infty}^{z_L} \frac{ds}{\sqrt{(s - \Re\{q\})^2 + \Im\{q\}^2}(s-z)} + 4 \int_{-\infty}^{z_L} \frac{ds}{\sqrt{\mu^2 - 16s}\sqrt{(s - \Re\{q\})^2 + \Im\{q\}^2}(s-z)} - 4 \int_0^{z^+} \frac{ds}{\sqrt{\mu^2 - 16s}\sqrt{(s - \Re\{q\})^2 + \Im\{q\}^2}(s-z)} + \frac{2}{\alpha} \int_{z^+}^{+\infty} \frac{ds}{\sqrt{(s - \Re\{q\})^2 + \Im\{q\}^2}(s-z)}.$$

This completes the extension of $Y(z; q, q^*)$ as an analytic function from z near \mathcal{B} to the maximal slit domain $\mathbb{C}_+ \cup \mathbb{C}_- \cup (z_L, 0)$.

The function $Y(z; q, q^*)$ has many purposes in our analysis. For example, it is easy to check with the use of elementary contour deformations that the following identities hold:

$$(88) \quad g'_+(z) - g'_-(z) = S_+(z)Y(z; q, q^*), \quad z \in \mathcal{B},$$

and

$$(89) \quad 2h'(z) = -S(z)Y(z; q, q^*), \quad z \in \mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty) \cup \mathcal{B} \cup \mathcal{B}^*).$$

The key property of the analytic function $Y(z; q, q^*)$ that we will require frequently is the following.

Proposition 1. *For $t > 0$ there exists exactly one simple zero of $Y(z; q, q^*)$ in its domain of analyticity (and hence by Schwarz symmetry necessarily located in the real interval $z_L < z < 0$). For $t < 0$ there exist exactly three zeros (counting with multiplicities) of $Y(z; q, q^*)$ in its domain of analyticity, and in this case given any $M > 0$, for $-t$ sufficiently small exactly one simple zero lies in the real interval $z_L < z < 0$ while there is an additional simple zero in each of the upper and lower half-planes with $|z| > M$. (In the boundary case of $t = 0$ the non-real zeros are at $z = \infty$.)*

Proof. The domain of analyticity of $Y(z; q, q^*)$ is the slit domain $z \in \mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty))$, and we note that Y as given by (87) satisfies the Schwarz symmetry condition $Y(z^*; q, q^*)^* = Y(z; q, q^*)$. From these facts and the Plemelj formula we find that upon taking a boundary value as z approaches the real axis from the upper half-plane,

$$(90) \quad \Im\{Y_+(z; q, q^*)\} = \begin{cases} \frac{4\pi\alpha - 2\pi\sqrt{\mu^2 - 16z}}{\alpha\sqrt{\mu^2 - 16z}\sqrt{(z - \Re\{q\})^2 + \Im\{q\}^2}}, & z < z_L \\ 0, & z_L < z < 0 \\ \frac{-4\pi}{\sqrt{\mu^2 - 16z}\sqrt{(z - \Re\{q\})^2 + \Im\{q\}^2}}, & 0 < z < z^+ \\ \frac{2\pi}{\alpha\sqrt{(z - \Re\{q\})^2 + \Im\{q\}^2}}, & z > z^+. \end{cases}$$

Also, from (87) it is clear that the complex boundary value $Y_+(z; q, q^*)$ is actually analytic except at the points $z \in \{z_L, 0, z^+\}$, where it becomes infinite in magnitude with logarithmic singularities. We may combine this information to show that, when the parameters correspond to the admissible region shown in Figure 1, the following statements are valid.

- (1) There is a positive constant $c > 0$ such that as z increases from $z = 0$ to $z = z^+$, $\Im\{Y_+(z; q, q^*)\} \leq -c < 0$, while $\Re\{Y_+(z; q, q^*)\} \rightarrow -\infty$ as $z \downarrow 0$ and $\Re\{Y_+(z; q, q^*)\} \rightarrow +\infty$ as $z \uparrow z^+$.
- (2) As z increases from $z = z^+$ to $z = +\infty$, $\Im\{Y_+(z; q, q^*)\} \geq 0$ with equality *only* in the limit $z \uparrow +\infty$, while $\Re\{Y_+(z; q, q^*)\} \rightarrow +\infty$ as $z \downarrow z^+$ and $\Re\{Y_+(z; q, q^*)\} \rightarrow 16t/\alpha^2$ as $z \uparrow +\infty$.
- (3) As z increases from $z = -\infty$ to $z = z_L$, $\Im\{Y_+(z; q, q^*)\} \leq 0$ with equality *only* in the limit $z \downarrow -\infty$, while $\Re\{Y_+(z; q, q^*)\} \rightarrow 16t/\alpha^2$ as $z \downarrow -\infty$ and $\Re\{Y_+(z; q, q^*)\} \rightarrow +\infty$ as $z \uparrow z_L$.

In particular, as z increases from $z = 0$ to $z = +\infty$ and then from $z = -\infty$ to $z = z_L$, the complex boundary value $Y_+(z; q, q^*)$ traces out a curve in the Y -plane that crosses the real axis only at one finite point, $Y = 16t/\alpha^2$, corresponding to $z = \pm\infty$. The three intervals in the z -plane and their images under Y_+ in the Y -plane are shown in red, with numbers corresponding to the above enumeration, in Figure 4. The

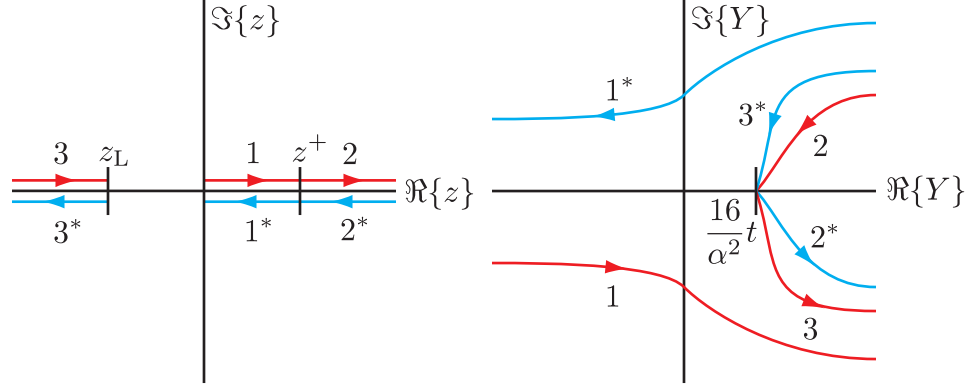


FIGURE 4. The domain of analyticity of $Y(z; q, q^*)$ is $\mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty))$, and a qualitative sketch of the image of the boundary of this slit domain in the Y -plane.

desired result then follows from an elementary application of the Argument Principle using the information enumerated above regarding the boundary value $Y_+(z; q, q^*)$ on the cut (and corresponding information regarding the boundary value $Y_-(z; q, q^*)$ follows by Schwarz symmetry). It is obvious that as the boundary of the slit domain is traversed once in the positive sense (in the order 1, 2, 3, 3*, 2*, and 1* as indicated in Figure 4), the image curve in the Y -plane encircles the origin exactly once in the positive sense for $t > 0$ and exactly three times in the positive sense for $t < 0$. See Figure 4. This gives the count of the zeros in the domain of analyticity. The proof is complete upon using continuity of Y with respect to t and Schwarz symmetry. \square

It is not difficult to obtain the asymptotic expansion of $Y(z; q, q^*)$ as $z \rightarrow \infty$ with $\Im\{z\} > 0$. The easiest way to do this is to analytically continue the representation (85) to z outside the loop L , by extracting a residue:

$$(91) \quad Y(z; q, q^*) = \left[\frac{1}{2\pi i} \oint_L \frac{\Phi'(s) ds}{S(s)(s-z)} + (*)^* \right] - \frac{\Phi'(z)}{S(z)} + \frac{16t}{\alpha^2}, \quad z \text{ outside of } L \text{ with } \Im\{z\} > 0.$$

Now we can let z tend to infinity by expanding the Cauchy kernel in a geometric series for L fixed. Thus:

$$(92) \quad \begin{aligned} \frac{1}{2\pi i} \oint_L \frac{\Phi'(s) ds}{S(s)(s-z)} + (*)^* &= -\Im \left\{ \frac{1}{\pi} \oint_L \frac{\Phi'(s) ds}{S(s)} \right\} \frac{1}{z} + O\left(\frac{1}{z^2}\right) \\ &= -\frac{2}{\pi} \Im \left\{ \int_B \frac{\Phi'(s) ds}{S_+(s)} \right\} \frac{1}{z} + O\left(\frac{1}{z^2}\right), \quad z \rightarrow \infty. \end{aligned}$$

Assuming that q is chosen so that at least the first of the equations (79) holds, we therefore find that

$$(93) \quad \frac{1}{2\pi i} \oint_L \frac{\Phi'(s) ds}{S(s)(s-z)} + (*)^* = \frac{4}{\alpha^2} (2(q + q^*)t + \alpha x - t) \frac{1}{z} + O\left(\frac{1}{z^2}\right), \quad z \rightarrow \infty.$$

The final step is to recall the expansions (54) and (108) to find

$$(94) \quad -\frac{\Phi'(z)}{S(z)} = \frac{2\pi i}{z} + O(z^{-3/2}), \quad z \rightarrow \infty, \quad \Im\{z\} > 0.$$

Using (93) with (94) in (91) then gives

$$(95) \quad Y(z; q, q^*) = \frac{16t}{\alpha^2} + \left[\frac{4}{\alpha^2} (2(q + q^*)t + \alpha x - t) + 2\pi i \right] \frac{1}{z} + O\left(\frac{1}{z^{3/2}}\right), \quad z \rightarrow \infty, \quad \Im\{z\} > 0.$$

4.3. The endpoint q as a function of x and t . Now we return to the conditions (79) supposed to be satisfied by the endpoint q of the arc \mathcal{B} , given $(x, t) \in \mathbb{R}^2$. First we dispense with the special case of $t = 0$.

Proposition 2. *The equations (79) hold for $t = 0$ and $x \leq x_c$ (for any arc \mathcal{B} connecting q to z_P lying in the quadrant given by the inequalities $\Im\{z\} > 0$ and $\Re\{z\} < z_P$) with $q = \mathfrak{z}(x)$ and $q^* = \mathfrak{z}(x)^*$ being the two branches of the complex part of the turning point curve.*

Proof. Set $t = 0$ and assume that $q = \mathfrak{z}(x)$. Whatever the arc \mathcal{B} that is the path of integration in both integrals in (79) actually is, it is homotopic (with orientation preserved) to the image of the map $s = \mathfrak{z}(y)$ as y decreases from x to $-\infty$, and the homotopy avoids the line of phantom poles $\Re\{s\} = z_P$. Since $\Phi'(s)$ is analytic for $\Im\{s\} > 0$ and since the boundary value $S_+(s)$ is analytic for $s \in \mathcal{B}$, we may deform the path of integration from \mathcal{B} to the image of $s = \mathfrak{z}(y)$. In the remainder of this proof, we suppose that this deformation is done, and that $S(z)$ has been redefined with its branch cut as this new contour and its Schwarz reflection in the lower half-plane.

Therefore, for $p = 0$ or $p = 1$, we may parametrize the contour integrals appearing in (79) by $-\infty < y < x$:

$$(96) \quad I_p(\mathfrak{z}(x), \mathfrak{z}(x)^*) = \Im \left\{ \int_{q=\mathfrak{z}(x)}^{z_P} \frac{\Phi'(s)s^p ds}{S_+(s)} \right\} = \Im \left\{ \int_{-\infty}^x \frac{\Phi'(\mathfrak{z}(y))\mathfrak{z}(y)^p \mathfrak{z}'(y) dy}{S_+(\mathfrak{z}(y))} \right\}.$$

(To get the sign correct it is important to take into account that $S_+(s)$ denotes the boundary value taken on the integration contour from the left according to the indicated direction of orientation; since S changes sign across its branch cuts the limits of integration in the s -integral can therefore be exchanged without changing the sign of the integral.) We will prove the proposition by showing that

$$(97) \quad I_0(\mathfrak{z}(x), \mathfrak{z}(x)^*) = -\frac{2\pi x}{\alpha} \quad \text{and} \quad I_1(\mathfrak{z}(x), \mathfrak{z}(x)^*) = -\frac{2\pi x}{\alpha} \Re\{q\}.$$

At this point (having used analyticity of $\Phi(s)$ apparent from the formula (48) obtained from Stirling asymptotics to deform contours) it becomes more convenient to use the alternate formula (61) for $\Phi(s)$ that is obtained from WKB theory. The WKB formula (61) is initially defined for real z for which there exist two turning points (in the present case this is the interval $z_L < z < 0$) but for the particular initial data under consideration it has an analytic continuation to $\Im\{z\} > 0$ that agrees exactly with the Stirling formula (48). By differentiation of $\Phi(s)$ as given by (61), we have

$$(98) \quad \Phi'(s) = \frac{1}{\alpha} \frac{d\omega_-}{ds}(s) x_-(s) + \frac{1}{\alpha} \int_{-\infty}^{x_-(s)} \left[\frac{\partial \omega}{\partial s}(v; s) - \frac{d\omega_-}{ds}(s) \right] dv, \quad z_L < s < 0,$$

where ω and ω_- are defined in (56)–(58), and where $x_-(s)$ is the real function defined for $z_L < s < 0$ by the portion of the turning point curve illustrated in red in Figure 3. Now, $x_-(s)$ admits analytic continuation into the complex s -plane from its interval of definition, and since it has a simple critical point at $s = z_c$ (a local minimum), there is a curve passing vertically through the point $s = z_c$ in the complex s -plane along which the analytic continuation of $x_-(s)$ is real and decreasing away from $s = z_c$. Clearly, the portion of this curve in the upper half s -plane coincides with the contour $s = \mathfrak{z}(y)$ for $-\infty < y < x_c$. Therefore, when $s = \mathfrak{z}(y)$, we have $x_-(s) = y$, where $x_-(s)$ denotes the analytic continuation of the similarly-named function from the interval $z_L < s < 0$. Using this information and substituting the analytic continuation of (98) along the path $s = \mathfrak{z}(y)$ for $-\infty < y < x < x_c$ into (96) we obtain

$$(99) \quad I_p(\mathfrak{z}(x), \mathfrak{z}(x)^*) = \Im \left\{ \frac{4}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} y dy + \frac{1}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} \int_{-\infty}^y \left[\frac{\partial \omega}{\partial s}(v; \mathfrak{z}(y)) - 4 \right] dv dy \right\},$$

where we have also used the fact that $d\omega_-(s)/ds = 4$ (independently of the functions $\rho(\cdot)$ and $u(\cdot)$).

We simplify the first term on the right-hand side of (99), writing it in terms of a double integral as follows:

$$(100) \quad \frac{4}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} y dy = \frac{4x}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} dy - \frac{4}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} \int_y^x dv dy.$$

Exchanging the order of integration in the double integral gives

$$(101) \quad \frac{4}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} y dy = \frac{4x}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} dy - \frac{4}{\alpha} \int_{-\infty}^x \int_{-\infty}^v \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} dy dv.$$

Reparametrizing the integrals over y by the map $s = \mathfrak{z}(y)$ this becomes

$$(102) \quad \frac{4}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} y dy = \frac{4x}{\alpha} \int_{z_P}^{q=\mathfrak{z}(x)} \frac{s^p ds}{S_+(s)} - \frac{4}{\alpha} \int_{-\infty}^x \int_{z_P}^{\mathfrak{z}(v)} \frac{s^p ds}{S_+(s)} dv.$$

Now we simplify the second term on the right-hand side of (99), by first exchanging the order of integration and then reparametrizing the inner integral by $s = \mathfrak{z}(y)$:

$$(103) \quad \begin{aligned} \frac{1}{\alpha} \int_{-\infty}^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} \int_{-\infty}^y \left[\frac{\partial \omega}{\partial s}(v; \mathfrak{z}(y)) - 4 \right] dv dy &= \frac{1}{\alpha} \int_{-\infty}^x \int_v^x \frac{\mathfrak{z}(y)^p \mathfrak{z}'(y)}{S_+(\mathfrak{z}(y))} \left[\frac{\partial \omega}{\partial s}(v; \mathfrak{z}(y)) - 4 \right] dy dv \\ &= \frac{1}{\alpha} \int_{-\infty}^x \int_{\mathfrak{z}(v)}^{q=\mathfrak{z}(x)} \frac{s^p}{S_+(s)} \left[\frac{\partial \omega}{\partial s}(v; s) - 4 \right] ds dv. \end{aligned}$$

Combining these results shows that

$$(104) \quad I_p(\mathfrak{z}(x), \mathfrak{z}(x)^*) = \Im \left\{ \frac{4x}{\alpha} \int_{z_P}^{q=\mathfrak{z}(x)} \frac{s^p ds}{S_+(s)} + \frac{1}{\alpha} \int_{-\infty}^x \left[\int_{\mathfrak{z}(v)}^{q=\mathfrak{z}(x)} \frac{\partial \omega}{\partial s}(v; s) \frac{s^p ds}{S_+(s)} - 4 \int_{z_P}^{q=\mathfrak{z}(x)} \frac{s^p ds}{S_+(s)} \right] dv \right\}.$$

(The integral over v cannot be broken up without sacrificing convergence.) Here the paths of integration in the s -integrals all lie along the image of the mapping $s = \mathfrak{z}(y)$, $-\infty < y < x_c$. Since $S(s^*) = S(s)^*$ and since for $v \in \mathbb{R}$ we have $\omega(v, s^*) = \omega(v, s)^*$ and hence $\partial \omega / \partial s(v, s^*) = \partial \omega / \partial s(v, s)^*$ where $\omega(v, s)$ is defined for complex s by analytic continuation from the interval $z_L < s < 0$, we easily obtain

$$(105) \quad I_p(\mathfrak{z}(x), \mathfrak{z}(x)^*) = \frac{2x}{i\alpha} J_p(x) + \frac{1}{2i\alpha} \int_{-\infty}^x \left[\int_{\mathfrak{z}(v)}^{\mathfrak{z}(x)} \frac{\partial \omega}{\partial s}(v; s) \frac{s^p ds}{S_+(s)} + \int_{\mathfrak{z}(x)^*}^{\mathfrak{z}(v)^*} \frac{\partial \omega}{\partial s}(v; s) \frac{s^p ds}{S_+(s)} - 4J_p(x) \right] dv.$$

where

$$(106) \quad J_p(x) := \int_{\mathfrak{z}(x)^*}^{\mathfrak{z}(x)} \frac{s^p ds}{S_+(s)}.$$

Here the paths of integration all lie along the image of the mapping $s = \mathfrak{z}(y)$, $-\infty < y < x_c$ and its Schwarz reflection (which in particular contains the branch cut of the re-defined function $S(s)$).

Now we evaluate all of the integrals over s by contour integration arguments (in particular, this will prove that the integrand of the v -integral in (105) vanishes identically). Since S changes sign across its branch cut, and since the latter connects $\mathfrak{z}(x)^*$ to $\mathfrak{z}(x)$, we have

$$(107) \quad J_p(x) = -\frac{1}{2} \oint \frac{s^p ds}{S(s)}$$

where the path of integration is a closed, positively-oriented loop that encloses the branch cut of S . We may now calculate $J_p(x)$ in terms of residues at $s = \infty$. Since

$$(108) \quad \frac{1}{S(s)} = \frac{1}{s} + \Re\{q\} \frac{1}{s^2} + O\left(\frac{1}{s^3}\right), \quad s \rightarrow \infty,$$

we see easily that

$$(109) \quad J_0(x) = -i\pi \quad \text{and} \quad J_1(x) = -i\pi \Re\{q\}.$$

Also, by implicit differentiation of the identity $\omega(v; s)^2 = 16\alpha^2 s \rho_0(v) + (4s - 1 + \alpha u_0(v))^2$ we easily obtain

$$(110) \quad \frac{\partial \omega}{\partial s}(v; s) = \frac{8\alpha^2 \rho_0(v) + 4(4s - 1 + \alpha u_0(v))}{\omega(v; s)}.$$

By definition of the complex part of the turning point curve, the roots of the quadratic $\omega(v; s)^2$ are exactly $s = \mathfrak{z}(v)$ and $s = \mathfrak{z}(v)^*$ for $v \in \mathbb{R}$ with $v < x_c$; thus $\omega(v; s)^2 = 16(s - \mathfrak{z}(v))(s - \mathfrak{z}(v)^*)$. To determine the function $\omega(v; s)$ as a function of s for fixed $v < x_c$ (that is, to determine the proper way to take the square root of $\omega(v; s)^2$), we proceed as follows. By definition, $\omega(v; s)$ is real and has the sign σ_- for $z_L < s < 0$, because v lies to the left of both real turning points $x_{\pm}(s)$ in this case. The value of σ_- is determined from (56) under the assumption that z lies in the interval in which there exist two real turning points: $z_L < z < 0$. But (56) can be written in the form

$$(111) \quad \sigma_- = \text{sgn}(4(z - z_P))$$

so since $z_P > 0$ we have $\sigma_- = -1$. Now since $v < x_c$, the roots of $\omega(v; s)^2$ form a complex-conjugate pair in the complex s -plane, and as $\omega(v; s)$ is a negative real function of s for $z_L < s < 0$ while $z_P > 0$, we may construct the analytic continuation from the interval $s \in (z_L, 0)$ to the complement of a branch cut connecting the two roots $\mathfrak{z}(v)$ and $\mathfrak{z}(v)^*$ coinciding with a sub-arc of the branch cut for $S(s)$, and normalized so that $\omega(v; s) = 4s + O(1)$ as $s \rightarrow \infty$. With the branch cut for $\omega(v; s)$ defined in this way, we can write $\omega(v; s)S_+(s) = 4\tilde{S}_+(v; s)$ where $\tilde{S}(v; s) := \omega(v; s)S(s)/4 = s^2 + O(s)$ as $s \rightarrow \infty$, and $\tilde{S}(s; v)$ is analytic except for *two* cuts: one connecting $\mathfrak{z}(v)$ to $q = \mathfrak{z}(x)$ and lying along the contour $s = \mathfrak{z}(y)$ for $-\infty < y < x_c$, and the other being the Schwarz reflection of the first one. At last we are in a position to evaluate the integral involving $\partial\omega/\partial s(v; s)$ by residues; we first use the fact that $\tilde{S}(v; s)$ changes sign across the contours of integration to write

$$(112) \quad \int_{\mathfrak{z}(v)}^{\mathfrak{z}(x)} \frac{\partial\omega}{\partial s}(v; s) \frac{s^p ds}{S_+(s)} + \int_{\mathfrak{z}(x)^*}^{\mathfrak{z}(v)^*} \frac{\partial\omega}{\partial s}(v; s) \frac{s^p ds}{S_+(s)} = -\frac{1}{2} \oint \frac{8\alpha^2 \rho_0(v) + 4(4s - 1 + \alpha u_0(v))}{4\tilde{S}(v; s)} s^p ds$$

where the path of integration is a closed, positively-oriented loop that encircles both of the branch cuts of $\tilde{S}(v; s)$. Now since

$$(113) \quad \frac{1}{\tilde{S}(v; s)} = \frac{1}{s^2} + \frac{\Re\{q\} + \Re\{\mathfrak{z}(v)\}}{s^3} + O\left(\frac{1}{s^4}\right), \quad s \rightarrow \infty,$$

a residue calculation shows that, for $p = 0$,

$$(114) \quad \int_{\mathfrak{z}(v)}^{\mathfrak{z}(x)} \frac{\partial\omega}{\partial s}(v; s) \frac{ds}{S_+(s)} + \int_{\mathfrak{z}(x)^*}^{\mathfrak{z}(v)^*} \frac{\partial\omega}{\partial s}(v; s) \frac{ds}{S_+(s)} = -4\pi i$$

and, for $p = 1$,

$$(115) \quad \int_{\mathfrak{z}(v)}^{\mathfrak{z}(x)} \frac{\partial\omega}{\partial s}(v; s) \frac{s ds}{S_+(s)} + \int_{\mathfrak{z}(x)^*}^{\mathfrak{z}(v)^*} \frac{\partial\omega}{\partial s}(v; s) \frac{s ds}{S_+(s)} = -4\pi i \left[\Re\{q\} + \Re\{\mathfrak{z}(v)\} + \frac{1}{2}\alpha^2 \rho_0(v) + \frac{\alpha u_0(v) - 1}{4} \right] \\ = -4\pi i \Re\{q\},$$

where on the second line we have used (78). Using (109), (114), and (115) in (105) we arrive at (97), so the proof is complete. \square

Now we show that the solution of the equations (79) can be continued to nearby (x, t) under certain conditions involving the function $Y(z; q, q^*)$.

Proposition 3. *Suppose that q is a solution of the equations (79) at $x = x_0$ and $t = t_0$ with $\Im\{q\} > 0$ and for which $Y(q; q, q^*) \neq 0$. Then the equations (79) have a unique smooth solution $q(x, t)$ valid for (x, t) near (x_0, t_0) for which $q(x_0, t_0) = q$.*

Proof. We construct the Jacobian determinant of M_0 and M_1 with respect to q and q^* (here viewed as independent complex variables). We first rewrite M_0 and M_1 in the form

$$(116) \quad M_0(q, q^*, x, t) = \frac{2\pi}{\alpha^2} (2(q + q^*)t + \alpha x - t) - \frac{i}{4} \left[\oint_L \frac{\Phi'(s) ds}{S(s)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)} \right] \\ M_1(q, q^*, x, t) = \frac{\pi}{\alpha^2} ((3q^2 + 2qq^* + 3q^{*2})t + (q + q^*)(\alpha x - t)) - \frac{i}{4} \left[\oint_L \frac{\Phi'(s)s ds}{S(s)} - \oint_{L^*} \frac{\Phi'(s)s ds}{S(s)} \right],$$

where L is a negatively-oriented contour beginning and ending at z_P and encircling the contour arc B exactly once, and L^* denotes the Schwarz reflection of this contour in the lower half-plane (with orientation induced

from L by reflection, that is, L^* is positively-oriented). Now holding L fixed, the only dependence in the integrals on q and q^* enters parametrically through the function $S(s)$. Since by direct calculation,

$$(117) \quad \frac{\partial}{\partial q} \frac{1}{S(s)} = \frac{1}{2S(s)(s-q)} \quad \text{and} \quad \frac{\partial}{\partial q^*} \frac{1}{S(s)} = \frac{1}{2S(s)(s-q^*)}$$

we find that

$$(118) \quad \begin{aligned} \frac{\partial M_0}{\partial q} &= \frac{4\pi t}{\alpha^2} - \frac{i}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)(s-q)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)(s-q)} \right] \\ \frac{\partial M_0}{\partial q^*} &= \frac{4\pi t}{\alpha^2} - \frac{i}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)(s-q^*)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)(s-q^*)} \right]. \end{aligned}$$

Comparing with the original formula (85) for $Y(z; q, q^*)$, we see that

$$(119) \quad \frac{\partial M_0}{\partial q} = \frac{\pi}{4} Y(q; q, q^*) \quad \text{and} \quad \frac{\partial M_0}{\partial q^*} = \frac{\pi}{4} Y(q^*; q, q^*).$$

Similarly, for partial derivatives of M_1 we obtain

$$(120) \quad \begin{aligned} \frac{\partial M_1}{\partial q} &= \frac{\pi}{\alpha^2} [6tq + 2tq^* + \alpha x - t] - \frac{i}{8} \left[\oint_L \frac{\Phi'(s)s ds}{S(s)(s-q)} - \oint_{L^*} \frac{\Phi'(s)s ds}{S(s)(s-q)} \right] \\ \frac{\partial M_1}{\partial q^*} &= \frac{\pi}{\alpha^2} [6tq^* + 2tq + \alpha x - t] - \frac{i}{8} \left[\oint_L \frac{\Phi'(s)s ds}{S(s)(s-q^*)} - \oint_{L^*} \frac{\Phi'(s)s ds}{S(s)(s-q^*)} \right]. \end{aligned}$$

The partial derivatives of M_1 can be rewritten as

$$(121) \quad \begin{aligned} \frac{\partial M_1}{\partial q} &= \frac{\pi}{\alpha^2} [6tq + 2tq^* + \alpha x - t] - \frac{iq}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)(s-q)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)(s-q)} \right] \\ &\quad - \frac{i}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)} \right] \\ \frac{\partial M_1}{\partial q^*} &= \frac{\pi}{\alpha^2} [6tq^* + 2tq + \alpha x - t] - \frac{iq^*}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)(s-q^*)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)(s-q^*)} \right] \\ &\quad - \frac{i}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)} \right]. \end{aligned}$$

Setting $x = x_0$ and $t = t_0$, we use the fact that q is a solution of $M_0 = 0$ to eliminate the terms on the second line in each case:

$$(122) \quad \begin{aligned} \left. \frac{\partial M_1}{\partial q} \right|_{(x_0, t_0)} &= \frac{4\pi t_0}{\alpha^2} q - \frac{iq}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)(s-q)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)(s-q)} \right] \\ &= q \left. \frac{\partial M_0}{\partial q} \right|_{(x_0, t_0)} \\ \left. \frac{\partial M_1}{\partial q^*} \right|_{(x_0, t_0)} &= \frac{4\pi t_0}{\alpha^2} q^* - \frac{iq^*}{8} \left[\oint_L \frac{\Phi'(s) ds}{S(s)(s-q^*)} - \oint_{L^*} \frac{\Phi'(s) ds}{S(s)(s-q^*)} \right] \\ &= q^* \left. \frac{\partial M_0}{\partial q^*} \right|_{(x_0, t_0)}. \end{aligned}$$

Therefore, the Jacobian determinant is

$$(123) \quad \begin{aligned} \mathcal{J}(q, q^*, x_0, t_0) &:= \left[\frac{\partial M_1}{\partial q} \frac{\partial M_0}{\partial q^*} - \frac{\partial M_0}{\partial q} \frac{\partial M_1}{\partial q^*} \right] \Big|_{(x_0, t_0)} = \frac{\pi^2}{16} Y(q; q, q^*) Y(q^*; q, q^*) (q - q^*) \\ &= \frac{i\pi^2}{8} |Y(q; q, q^*)|^2 \Im\{q\}. \end{aligned}$$

Since this is nonzero by assumption, the solution to the simultaneous equations $M_0 = M_1 = 0$ can be continued from q to $q(x, t)$ by the Implicit Function Theorem. \square

Corollary 1. *Let $x < x_c$. Then, for $|t|$ sufficiently small, there exists a unique solution $q = q(x, t)$ of the equations (79) satisfying $q(x, 0) = \mathfrak{z}(x)$, where $\mathfrak{z}(x)$ is the complex part of the turning point curve. For $t > 0$ the solution can be continued (even for $x \geq x_c$) until $\Im\{q(x, t)\}$ becomes zero.*

Proof. From Proposition 2 we have that the equations (79) are satisfied by $q = \mathfrak{z}(x)$ when $t = 0$ and $x < x_c$, and we note that $\Im\{\mathfrak{z}(x)\} > 0$ for $x < x_c$. By Proposition 1 the only possible zeros of $Y(z; q, q^*)$ in the open upper half-plane are far from q when $|t|$ is small, which implies that $Y(\mathfrak{z}(x); \mathfrak{z}(x), \mathfrak{z}(x)^*) \neq 0$. Moreover, if $t > 0$ then $Y(q; q, q^*)$ can only become zero if $\Im\{q\} = 0$. Therefore, by Proposition 3 the Corollary is proved. \square

It follows that for $t \geq 0$, the only obstruction to smooth continuation of the endpoint function $q(x, t)$ from its values $q(x, 0) = \mathfrak{z}(x)$ for $x < x_c$ may be a collision of q with the real axis. We now completely resolve the scope of the continuation.

Proposition 4. *There exists a well-defined smooth curve $x = x_c(t)$ for $t \geq 0$ satisfying $x_c(0) = x_c$, such that the solution of (79) with $q(x, 0) = \mathfrak{z}(x)$ for $x < x_c$ can be uniquely continued to the domain $t > 0$ and $x < x_c(t)$. Also, $\Im\{q(x, t)\} > 0$ whenever $x < x_c(t)$, but $\Im\{q(x, t)\} \downarrow 0$ as $x \uparrow x_c(t)$.*

Proof. First, we argue that it is impossible for $\Im\{q\}$ to vanish for bounded (x, t) unless q tends to a real point z in the interval $(z_L, 0)$. Consider the combination of M_0 and M_1 given by:

$$(124) \quad \begin{aligned} F(q, q^*; x, t) &:= M_1(q, q^*; x, t) - \frac{1}{2}(q + q^*)M_0(q, q^*; x, t) \\ &= \frac{\pi t}{\alpha^2}(q - q^*)^2 + I_1(q, q^*) - \frac{1}{2}(q + q^*)I_0(q, q^*). \end{aligned}$$

Clearly, we must have $F(q, q^*; x, t) = 0$ for any solution of the two equations $M_0 = M_1 = 0$. Now for t bounded the first term on the second line above obviously converges to zero as $\Im\{q\} \downarrow 0$. By definition of $I_p(q, q^*)$ we have

$$(125) \quad I_1(q, q^*) - \frac{1}{2}(q + q^*)I_0(q, q^*) = \Im \left\{ \int_B \Phi'(s) \frac{s - \frac{1}{2}(q + q^*)}{S_+(s)} ds \right\}.$$

A careful dominated convergence argument (in which the integration from q is replaced locally by a loop around $s = q$ that is deformed to the real axis in a neighborhood of $z \in \mathbb{R}$, a contour on which $(s - \frac{1}{2}(q + q^*)) / S(s)$ converges pointwise to 1 and is uniformly bounded) shows that

$$(126) \quad \lim_{\substack{q \rightarrow z \in \mathbb{R} \\ \Im\{q\} > 0}} \left[I_1(q, q^*) - \frac{1}{2}(q + q^*)I_0(q, q^*) \right] = \Im \left\{ \int_z^{z_P} \Phi'_+(s) ds \right\} = -\Im\{\Phi_+(z)\}$$

because $\Im\{\Phi_+(z_P)\} = 0$ according to (53). But again according to (53) we see that $\Im\{\Phi_+(z)\} = 0$ only for $z = z_P$ and for $z \in [z_L, 0]$. Therefore, for bounded t we can only have $\Im\{q\} \downarrow 0$ if $q \rightarrow z \in [z_L, 0]$ or if $q \rightarrow z_P$. But a similar argument also shows that $M_0(q, q^*; x, t) \rightarrow \infty$ if $q \rightarrow z_L$, $q \rightarrow 0$, or if $q \rightarrow z_P$ with $\Im\{q\} > 0$ and (x, t) bounded, so q can only tend to a point $z \in (z_L, 0)$ for bounded (x, t) .

Now fix $z \in (z_L, 0)$. Then $F(q, q^*; x, t) \rightarrow 0$ as $q \rightarrow z$ with $\Im\{q\} > 0$, so the two conditions $M_0 = M_1 = 0$ degenerate to a single condition: $M_0^c(z; x, t) = 0$, where

$$(127) \quad M_0^c(z; x, t) := \frac{2\pi}{\alpha^2}(4zt + \alpha x - t) + I_0^c(z), \quad z_L < z < 0,$$

where

$$(128) \quad \begin{aligned} I_0^c(z) &:= -\frac{\pi}{2}\Phi'(z) + \Im \left\{ \int_0^{z_P} \frac{\Phi'_+(s) ds}{s - z} \right\} \\ &= \frac{\pi}{2}\Phi'(z) + \Im \left\{ \int_{z_L}^{z_P} \frac{\Phi'(s) ds}{s - z} \right\}, \end{aligned}$$

where on the second line the path of integration from z_L to z_P lies in the upper half s -plane. Since $\Phi'(z) \rightarrow +\infty$ as $z \downarrow z_L$ while $\Phi'(z) \rightarrow -\infty$ as $z \uparrow 0$, the expression on the first line shows that $I_0^c(z) \rightarrow -\infty$ as $z \downarrow z_L$, while that on the second line shows that $I_0^c(z) \rightarrow -\infty$ as $z \uparrow 0$. Both expressions are real-valued, finite, and equivalent for $z \in (z_L, 0)$.

For each $z \in (z_L, 0)$, the condition $M_0^c(z; x, t) = 0$ defines a straight line in the (x, t) -plane. We will now show that the union of these lines is an unbounded region of the form $x \geq x_c(t)$ for all $t \geq 0$, where $x_c(t)$ is a smooth function of t with $x_c(0) = x_c$. More precisely, through each point (x, t) with $x > x_c(t)$ and $t \geq 0$ there will pass exactly two of these lines corresponding to distinct values of $z \in (z_L, 0)$, and the curve $x = x_c(t)$ will appear as the caustic formed by the intersection of infinitesimally neighboring lines. Given $t \geq 0$ fixed, the possible x -values lying on the lines are given by

$$(129) \quad x = \frac{1-4z}{\alpha}t - \frac{\alpha}{2\pi}I_0^c(z), \quad z_L < z < 0.$$

The first term is obviously bounded for t fixed and $z \in (z_L, 0)$, but the second term tends to $+\infty$ as $z \downarrow z_L$ or $z \uparrow 0$. This shows that for $t \geq 0$ fixed the region occupied by the lines is exactly

$$(130) \quad x \geq x_c(t) := \min_{z_L < z < 0} \left[\frac{1-4z}{\alpha}t - \frac{\alpha}{2\pi}I_0^c(z) \right] < +\infty.$$

To determine $x_c(t)$ we show that for each $t \geq 0$, the function to be minimized has exactly one critical point in $(z_L, 0)$, which is then (by asymptotics as $z \downarrow z_L$ and as $z \uparrow 0$) necessarily the unique minimizer. Indeed,

$$(131) \quad \frac{\partial}{\partial z} \left[\frac{1-4z}{\alpha}t - \frac{\alpha}{2\pi}I_0^c(z) \right] = -\frac{\alpha}{2\pi} \frac{\partial M_0^c}{\partial z}(z; x, t) = -\frac{\alpha}{4}Y^c(z; z),$$

where $Y^c(w; z)$ is by definition for $z_L < w < 0$ the limiting value of the function $Y(w; q, q^*)$ as $q \rightarrow z \in (z_L, 0)$ with $\Im\{q\} > 0$:

$$(132) \quad Y^c(w; z) := \frac{16t}{\alpha^2} - \frac{\Phi'(w) - \Phi'(z)}{w - z} + \frac{2}{\pi} \Im \left\{ \int_0^{z_P} \frac{\Phi'_+(s) ds}{(s-z)(s-w)} \right\}, \quad z_L < w < 0, \quad z_L < z < 0.$$

It is obvious that $Y^c(w; z)$ is symmetric in its arguments: $Y^c(w; z) = Y^c(z; w)$, and for each $z \in (z_L, 0)$ there is, by Proposition 1, a unique simple zero $w = w(z) \in (z_L, 0)$ of $Y^c(w; z)$. The graph $w = w(z)$ must be symmetric with respect to reflection through the diagonal $w = z$ because otherwise there would necessarily be multiple roots $w \in (z_L, 0)$ of $Y^c(w; z)$ for at least some z in the interval $(z_L, 0)$. Therefore, this graph either coincides with the diagonal line $w = z$ or it connects $(z_L, 0)$ with $(0, z_L)$ and crosses the diagonal exactly once orthogonally. But the case $w(z) = z$ can be ruled out easily by symmetry of $Y^c(w; z)$ and the fact that $Y^c(w; z) \uparrow +\infty$ as $w \downarrow z_L$ while $Y^c(w; z) \downarrow -\infty$ as $w \uparrow 0$. Therefore, $Y^c(z; z)$ has exactly one simple root in the interval $(z_L, 0)$ for all $t > -T$, and this establishes the uniqueness of the critical point and hence of the minimizer. Let us denote this critical point by $z = z_c(t)$, and the corresponding value of x defined for $z = z_c(t)$ by (129) as $x = x_c(t)$. By consideration of a different type of configuration for the g -function in which the branch points of the square root $S(z)$ are both real, and examination of the limit in which these two points coalesce (see [9]), it can be shown that in fact $z_c(0) = z_c \in (z_L, 0)$, where z_c is the specific value at which the complex branches $\mathfrak{z}(x)$ and $\mathfrak{z}(x)^*$ of the turning point curve become real (see Figure 3). It then follows from Proposition 2 that $x_c(0) = x_c$. \square

The curve $x = x_c(t)$ is plotted for the case of $\alpha = \delta = 1$ and $\mu = 2$ in Figure 5. In Figure 6 we plot the differences between $x_c(t)$ for various parameter values and the specific function $x_c(t)$ shown in Figure 5. In particular, these plots clearly show that $x_c(t)$ is not (in general at least) a linear function of t .

Finally, we show that the complex-valued function $q(x, t)$ defined for all $t \geq 0$ and $x \leq x_c(t)$ satisfies certain canonical partial differential equations (Whitham equations) in the interior of the domain of definition.

Proposition 5. *Suppose the equations (79) admit a solution $q = q(x, t)$ differentiable with respect to x and t . Then*

$$(133) \quad \frac{\partial q}{\partial t} + \frac{1}{\alpha} [1 - 3q - q^*] \frac{\partial q}{\partial x} = 0 \quad \text{and} \quad \frac{\partial q^*}{\partial t} + \frac{1}{\alpha} [1 - 3q^* - q] \frac{\partial q^*}{\partial x} = 0.$$

Proof. This follows from the relations $M_0(q(x, t), q^*(x, t), x, t) = M_1(q(x, t), q^*(x, t), x, t) = 0$ by implicit differentiation. Indeed, since

$$(134) \quad \frac{\partial M_0}{\partial x} = \frac{2\pi}{\alpha} \quad \text{and} \quad \frac{\partial M_0}{\partial t} = \frac{2\pi}{\alpha^2} [2(q + q^*) - 1]$$

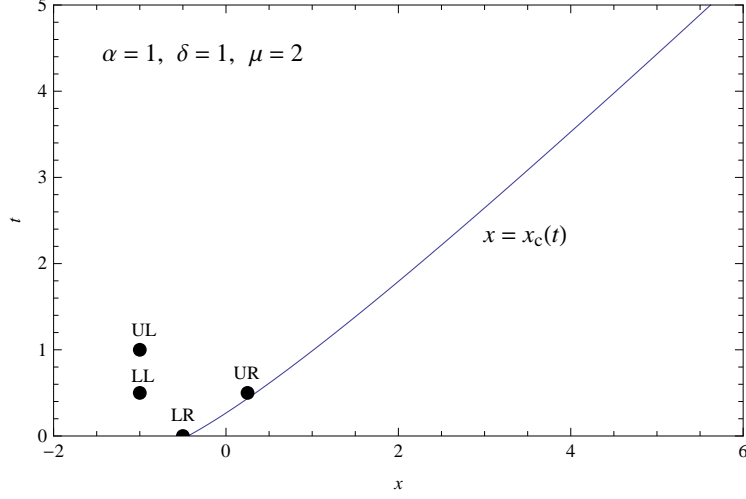


FIGURE 5. The curve $x_c(t)$ for $\alpha = \delta = 1$ and $\mu = 2$. The four marked points in the domain $x < x_c(t)$ correspond to the (x, t) -values in the four indicated panels of Figure 7 below.

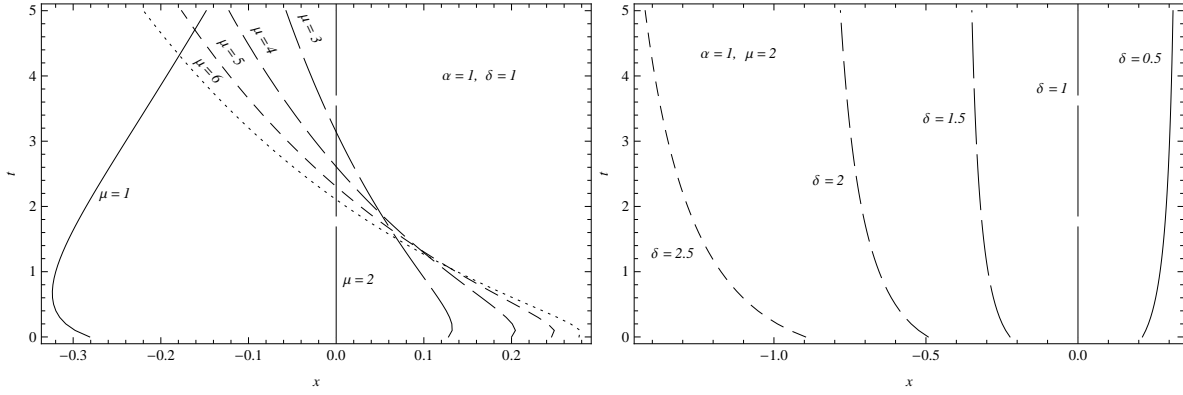


FIGURE 6. Left: for $\alpha = \delta = 1$, curves $x = x_c(t)$ for various values of μ compared with the curve in Figure 5. Right: for $\alpha = 1$ and $\mu = 2$, curves $x = x_c(t)$ for various values of δ compared with the curve in Figure 5.

while

$$(135) \quad \frac{\partial M_1}{\partial x} = \frac{\pi}{\alpha}(q + q^*) \quad \text{and} \quad \frac{\partial M_1}{\partial t} = \frac{\pi}{\alpha^2}[3q^2 + 2qq^* + 3q^{*2} - q - q^*],$$

it follows immediately from (119) and (122) given in the proof of Proposition 3 that

$$(136) \quad \frac{\partial q}{\partial x} = -\frac{4}{\alpha Y(q; q, q^*)} \quad \text{and} \quad \frac{\partial q^*}{\partial x} = -\frac{4}{\alpha Y(q^*; q, q^*)},$$

and also that

$$(137) \quad \frac{\partial q}{\partial t} = \frac{4}{\alpha^2 Y(q; q, q^*)} [1 - 3q - q^*] \quad \text{and} \quad \frac{\partial q^*}{\partial t} = \frac{4}{\alpha^2 Y(q^*; q, q^*)} [1 - 3q^* - q].$$

It is then obvious that (133) holds true if $Y(q; q, q^*)$ and $Y(q^*; q, q^*)$ are nonzero, a condition for the existence of a smooth solution $q(x, t)$. \square

The equations (133) are actually equivalent to the dispersionless MNLS system (9) in the locally subsonic case that $Q < 0$, as can be seen by means of a straightforward calculation using the substitutions

$$(138) \quad \rho = \frac{4}{\alpha^2} (\Im\{q^{1/2}\})^2 \quad \text{and} \quad u = \frac{1}{\alpha} (1 - 4|q|).$$

This shows that the variables q and q^* are complex *Riemann invariants* for the elliptic dispersionless MNLS system.

4.4. Choice of the arc \mathcal{B} . Supposing that $t \geq 0$ and $x < x_c(t)$, we determine the endpoint $q = q(x, t)$ with $\Im\{q(x, t)\} > 0$ as explained in §4.3 and hence obtain as explained in §4.2 the functions $g(z)$ and $h(z)$ (parametrized by (x, t) of course). These functions are actually only well-defined once we specify a particular contour arc \mathcal{B} in the upper half-plane connecting $z = q$ with $z = z_P$. We now describe how \mathcal{B} is to be chosen.

Proposition 6. *Let $t \geq 0$ and $x < x_c(t)$, and let $q = q(x, t)$ be determined as explained in §4.3. Then there is a unique choice of the arc \mathcal{B} connecting $z = q$ with $z = z_P$ for which $\Im\{h(z)\}$ can be extended by continuity to the domain $z \in \mathbb{C} \setminus (-\infty, z_L] \cup [0, +\infty)$.*

Proof. The key observation is that as the imaginary part of an analytic function, $\Im\{h(z)\}$ is a harmonic function in the domain $\mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty) \cup \mathcal{B} \cup \mathcal{B}^*)$. Generally there is a jump discontinuity across $\mathcal{B} \cup \mathcal{B}^*$ that we wish to avoid by choice of \mathcal{B} .

Firstly, regardless of how \mathcal{B} is chosen, we can easily see that $\Im\{h(z)\} = 0$ for $z \in [z_L, 0]$, for $z = z_P$, for $z = q$, and for $z = q^*$. Indeed, by Schwarz symmetry of h and the fact that $(z_L, 0)$ is in the domain of analyticity of h , it follows immediately that $\Im\{h(z)\} = 0$ for $z_L < z < 0$. But since $S(z)$ is Schwarz symmetric and analytic on the real axis except at its jump discontinuity point $z = z_P$, and since according to (90), the boundary values $\Im\{Y_{\pm}(z; q, q^*)\}$ taken by $\Im\{Y(z; q, q^*)\}$ on the real axis are bounded, it follows from (89) that the corresponding boundary values of $\Im\{h(z)\}$ are real differentiable functions of real z except at the point $z = z_P$ (which is nonetheless a point of Lipschitz continuity of the boundary values of $\Im\{h(z)\}$ for $z \in \mathbb{R}$). In particular this implies that the limit points $z = z_L$ and $z = 0$ of the open interval $(z_L, 0)$ are also points where $\Im\{h(z)\} = 0$. The fact that the well-defined value $\Im\{h(z_P)\}$ vanishes is a consequence of the Fundamental Theorem of Calculus, the formula (89), and the formula (90); since $S(z) = -\sqrt{(z - \Re\{q\})^2 + \Im\{q\}^2}$ holds for $z < z_P$, we integrate along the top edge of the branch cut for $Y(z; q, q^*)$ to obtain

$$(139) \quad \begin{aligned} \Im\{h(z_P)\} &= \Im \left\{ h(0) - \frac{1}{2} \int_0^{z_P} S(s) Y_+(s; q, q^*) ds \right\} \\ &= \frac{1}{2} \int_0^{z_P} \sqrt{(s - \Re\{q\})^2 + \Im\{q\}^2} \Im\{Y_+(s; q, q^*)\} ds \\ &= -2\pi \int_0^{z^+} \frac{ds}{\sqrt{\mu^2 - 16s}} + \frac{\pi}{\alpha} \int_{z^+}^{z_P} ds. \end{aligned}$$

By direct evaluation of these integrals and the use of the definitions of z^+ and z_P we obtain the claimed result that $\Im\{h(z_P)\} = 0$. By Schwarz symmetry of $Y(z; q, q^*)$ the same result holds had we integrated instead along the lower edge of the branch cut for $Y(z; q, q^*)$. With this information we can express $\Im\{h(q)\}$ as an integral again using the Fundamental Theorem of Calculus. Indeed, integrating along the left edge of the branch cut \mathcal{B} in the direction of its orientation from $z = q$ to $z = z_P$, we get (using $\Im\{h(z_P)\} = 0$)

$$(140) \quad \Im\{h(q)\} = \frac{1}{2} \Im \left\{ \int_{\mathcal{B}} S_+(s) Y(s; q, q^*) ds \right\}.$$

Now using the formula (88) and the fact that $g(z)$ is Schwarz-symmetric and analytic for $z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*)$ we have

$$(141) \quad \Im\{h(q)\} = \frac{1}{2} \Im \left\{ \int_{\mathcal{B}} (g'_+(s) - g'_-(s)) ds \right\} = \frac{i}{4} \oint g'(s) ds = \frac{i}{4} \oint p(s) ds,$$

where the closed contour of integration is a large positively-oriented circle of arbitrarily large radius. It therefore follows from the fact that q satisfies the conditions (79) that $p(s) = O(s^{-2})$ as $s \rightarrow \infty$ and hence $\Im\{h(q)\} = 0$ as claimed. The fact that $\Im\{h(q^*)\} = 0$ then follows by Schwarz symmetry of h .

Still regarding the arc \mathcal{B} connecting $z = q$ with $z = z_P$ as arbitrary, we note that the zero level set of $\Im\{h(z)\}$ defined by $\mathcal{Z} := \{z \in \mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty) \cup \mathcal{B} \cup \mathcal{B}^*), \Im\{h(z)\} = 0\}$ actually extends by continuity to the excluded branch cut $\mathcal{B} \cup \mathcal{B}^*$ and moreover is independent of \mathcal{B} . This follows from the fact that given $\Im\{h(q)\} = 0$ we may express $\Im\{h(z)\}$ in terms of an integral from $s = q$ to $s = z$, and because $s = q$ is the square-root branch point of $S(s)$, a change of choice of branch for S simply amounts to a change of sign of the integral and hence of $\Im\{h(z)\}$. This clearly leaves \mathcal{Z} invariant.

We will now characterize \mathcal{Z} completely, and prove in particular that \mathcal{Z} contains a smooth arc connecting $z = q$ with $z = z_P$. When the proof is finished we will select this arc to be the arc \mathcal{B} and show that this choice makes $\Im\{h(z)\}$ continuous. We firstly characterize the real points of \mathcal{Z} . We already know that \mathcal{Z} contains the real interval $[z_L, 0]$ and the real point z_P . Moreover, it is easy to see that these points exhaust $\mathcal{Z} \cap \mathbb{R}$. Indeed, since $S(z)$ is bounded away from zero for real $z \neq z_P$ and $\text{sgn}(S(z)) = \text{sgn}(z - z_P)$, it follows from (89) and (90) (see also Figure 4) that $\Im\{h_+(z)\}$ is strictly increasing for $z^+ < z < z_P$ and strictly decreasing for $z < z_L$, for $0 < z < z^+$, and for $z > z_P$.

Next we consider the points of \mathcal{Z} in the upper half-plane near the real axis. According to Proposition 1, for $t \geq 0$ there exists a unique simple zero $z = \xi$ of $Y(z; q, q^*)$ in the interval $(z_L, 0)$, and since $S(z)$ is nonzero near this point, $z = \xi$ is obviously a simple saddle point of $\Im\{h(z)\}$ and therefore there is a unique branch of \mathcal{Z} emanating from this point with a vertical tangent into the upper half-plane. Also, since the real derivative of the boundary value $\Im\{h_+(z)\}$ has a jump discontinuity at $z = z_P$ with opposite nonzero left and right limits, there is a unique branch of \mathcal{Z} emanating transversely to the real axis into the upper half-plane from $z = z_P$ (here the tangent is not necessarily vertical, however). The points $z = \xi \in (z_L, 0)$ and $z = z_P > 0$ are the only real limit points of the part of \mathcal{Z} in the open upper half-plane.

Next, since $z = q$ is a simple root of $S(z)^2$, a local analysis of $S(z)Y(z; q, q^*)$ near $z = q$ using the fact (see Proposition 1) that $Y(q; q, q^*) \neq 0$ shows that there exist exactly three arcs of \mathcal{Z} emanating from $z = q \in \mathcal{Z}$ separated by angles of $2\pi/3$.

Let us now analyze \mathcal{Z} assuming that $|z|$ is large with $\Im\{z\} > 0$. Since $S(s) = s[1 - \frac{1}{2}(q + q^*)s^{-1} + O(s^{-2})]$ as $s \rightarrow \infty$, we recall (89) and (95) to obtain

$$\begin{aligned} \Im\{h(z)\} &= -\frac{1}{2}\Im\left\{\int_q^z S(s)Y(s; q, q^*) ds\right\} \\ (142) \quad &= \Im\left\{-\frac{4t}{\alpha^2}z^2 - \left(\frac{2}{\alpha^2}(\alpha x - t) + \pi i\right)z + O(z^{1/2})\right\}, \quad z \rightarrow \infty, \quad \Im\{z\} > 0. \end{aligned}$$

For $t = 0$, the part of the level set \mathcal{Z} in the distant upper half-plane is clearly a smooth curve asymptotic to the straight line through the points $z = 0$ and $z = -2x/\alpha + \pi i$ as $z \rightarrow \infty$. For $t \neq 0$ however, \mathcal{Z} is asymptotic to leading order to the union of the real and imaginary axes, and the computation of higher order corrections is necessary to determine whether the horizontal asymptotes actually correspond to branches of \mathcal{Z} in the upper half-plane. Representing a branch of \mathcal{Z} in the form $z = |z|e^{i\phi}$, $\phi = \phi(|z|)$, we substitute into (142) and set $\Im\{h(z)\} = 0$ to obtain the relation

$$(143) \quad -\frac{4t}{\alpha^2}\sin(2\phi) - \frac{2}{\alpha^2|z|}(\alpha x - t)\sin(\phi) - \frac{\pi}{|z|}\cos(\phi) + O(|z|^{-3/2}) = 0, \quad |z| \rightarrow \infty.$$

Setting $|z| = \infty$ gives $\sin(2\phi) = 0$, all roots $\phi = \phi_0$ of which are simple. We may therefore apply the Implicit Function Theorem to continue these roots to finite $|z|$, yielding the asymptotic expansion

$$(144) \quad \phi(|z|) = \phi_0 - \frac{(\alpha x - t)}{4t|z|}\frac{\sin(\phi_0)}{\cos(2\phi_0)} - \frac{\pi\alpha^2}{8t|z|}\frac{\cos(\phi_0)}{\cos(2\phi_0)} + O(|z|^{-3/2}), \quad |z| \rightarrow \infty,$$

where ϕ_0 is a root of $\sin(2\phi_0) = 0$. Since we are only concerned with the upper half z -plane, we need to consider only the angles $\phi_0 = 0, \frac{1}{2}\pi, \pi$. For $\phi_0 = \frac{1}{2}\pi$, the asymptotic formula predicts a slight deformation of the angle in a direction depending on x and t , but regardless this branch remains in the upper half-plane due to the dominant constant term in the angle ϕ . For $\phi_0 = 0$ and $\phi_0 = \pi$, we have $\sin(\phi_0) = 0$, and the surviving term at order $|z|^{-1}$ in the asymptotic expansion (144) indicates that for $t > 0$ we have $|\frac{1}{2}\pi - \phi(|z|)| > 0$ for large $|z|$, showing that these solutions *do not* correspond to branches of \mathcal{Z} . Therefore, we conclude that for $t \geq 0$ the part of \mathcal{Z} in the distant upper half plane consists of exactly one smooth curve tending to infinity asymptotic to a non-horizontal straight line. (If $t < 0$ then the same analysis shows that when $\phi_0 = 0$ or

$\phi_0 = \pi$ we have instead $|\frac{1}{2}\pi - \phi(|z|)| < 0$ for large $|z|$ and there are then three distinct branches of \mathcal{Z} in the distant upper half-plane.)

So, we know that in the open upper half-plane there exist branches of \mathcal{Z} emanating from the real axis only from $z = \xi \in (z_L, 0)$ and from $z = z_P$. There also exist exactly three branches emanating from $z = q$, and (for $t \geq 0$) there is exactly one unbounded branch in the upper half-plane. Since $\Im\{h(z)\}$ is harmonic for $\Im\{z\} > 0$ except along the branch cut \mathcal{B} across which $\Im\{h(z)\}$ simply changes sign, the Maximum Principle prohibits the existence of any isolated components of \mathcal{Z} in the open upper half-plane, with the possible exception of curves enclosing $z = q$. But the latter are also easily excluded from \mathcal{Z} by local arguments. Therefore, the part of \mathcal{Z} in the upper half-plane consists exactly of the six branches described above, and these must be matched with each other. The Maximum Principle also prevents any of the arcs emanating from $z = q$ from coinciding with one another. Finally, since Proposition 1 rules out any zeros of $Y(z; q, q^*)$ in the open upper half-plane for $t \geq 0$, there are no saddle points of $\Im\{h(z)\}$ for $\Im\{z\} > 0$ and this means that the six arcs must be matched *pairwise*, and no crossings are allowed. Since none of the three branches emanating from $z = q$ can be matched with each other, we conclude that exactly one of these three branches is connected to each of the three distinct points $z = \xi$, $z = z_P$, and $z = \infty$. As the zero level set is Schwarz-symmetric, this concludes the complete description of \mathcal{Z} in the case $t \geq 0$.

Now suppose that the branch cut \mathcal{B} is taken to coincide with the branch (a smooth arc, actually) of \mathcal{Z} connecting $z = q$ with $z = z_P$. It then follows that $\Im\{h(z)\}$ can be continuously extended to $z \in \mathcal{B} \cup \mathcal{B}^*$. Indeed, since $\Im\{h(q)\} = 0$, from (89) and the Fundamental Theorem of Calculus, the mismatch $\Im\{h_+(z)\} - \Im\{h_-(z)\}$ of boundary values taken by $\Im\{h(z)\}$ on \mathcal{B} is

$$\begin{aligned} \Im\{h_+(z)\} - \Im\{h_-(z)\} &= -\frac{1}{2}\Im\left\{\int_q^z S_+(s)Y(s; q, q^*)ds - \int_q^z S_-(s)Y(s; q, q^*)ds\right\} \\ (145) \qquad &= -\Im\left\{\int_q^z S_+(s)Y(s; q, q^*)ds\right\}, \quad z \in \mathcal{B}. \end{aligned}$$

But this is the same as $2\Im\{h_+(z)\}$, which vanishes because $z \in \mathcal{B} \subset \mathcal{Z}$. \square

From now on we assume that for all $t \geq 0$ and $x < x_c(t)$, the arc \mathcal{B} is chosen exactly as explained above, making $\Im\{h(z)\}$ continuous across $\mathcal{B} \cup \mathcal{B}^*$. The zero level set \mathcal{Z} clearly divides the open upper half-plane into three disjoint open regions. We denote by \mathcal{V} (the *valley* of $\Im\{h(z)\}$ in the upper half-plane) the union of the two open regions that are separated by \mathcal{B} , and we denote by \mathcal{M} (the *mountain* of $\Im\{h(z)\}$ in the upper half-plane) the remaining open region. According to (142) and the fact that \mathcal{Z} contains all of the points for which $\Im\{h(z)\} = 0$, we have the following strict inequalities for $\Im\{z\} > 0$:

$$(146) \qquad \Im\{h(z)\} > 0, \quad z \in \mathcal{M}$$

and

$$(147) \qquad \Im\{h(z)\} < 0, \quad z \in \mathcal{V}.$$

The regions \mathcal{M} and \mathcal{V} are illustrated for several choices (x, t) in Figure 7.

4.5. Key properties of $g(z)$ and $h(z)$. With $g(z)$ and $h(z)$ now completely determined as analytic functions depending smoothly on parameters (x, t) with $t \geq 0$ and $x < x_c(t)$, we now explain all of the important consequences of our finished construction.

Proposition 7. *The function g is analytic for $z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*)$, satisfies the Schwarz symmetry condition $g(z^*) = g(z)^*$, and also $g(0) = 0$ while $g(\infty)$ is well-defined and finite. The function h is analytic for $z \in \mathbb{C} \setminus ((-\infty, z_L] \cup [0, +\infty) \cup \mathcal{B} \cup \mathcal{B}^*)$, satisfies the Schwarz symmetry condition $h(z^*) = h(z)^*$, and $\Im\{h(z)\}$ is continuous for $\Im\{z\} \neq 0$. There is a real constant $\kappa = \kappa(x, t)$ such that*

$$(148) \qquad h_+(z) + h_-(z) = \kappa, \quad z \in \mathcal{B} \cup \mathcal{B}^*.$$

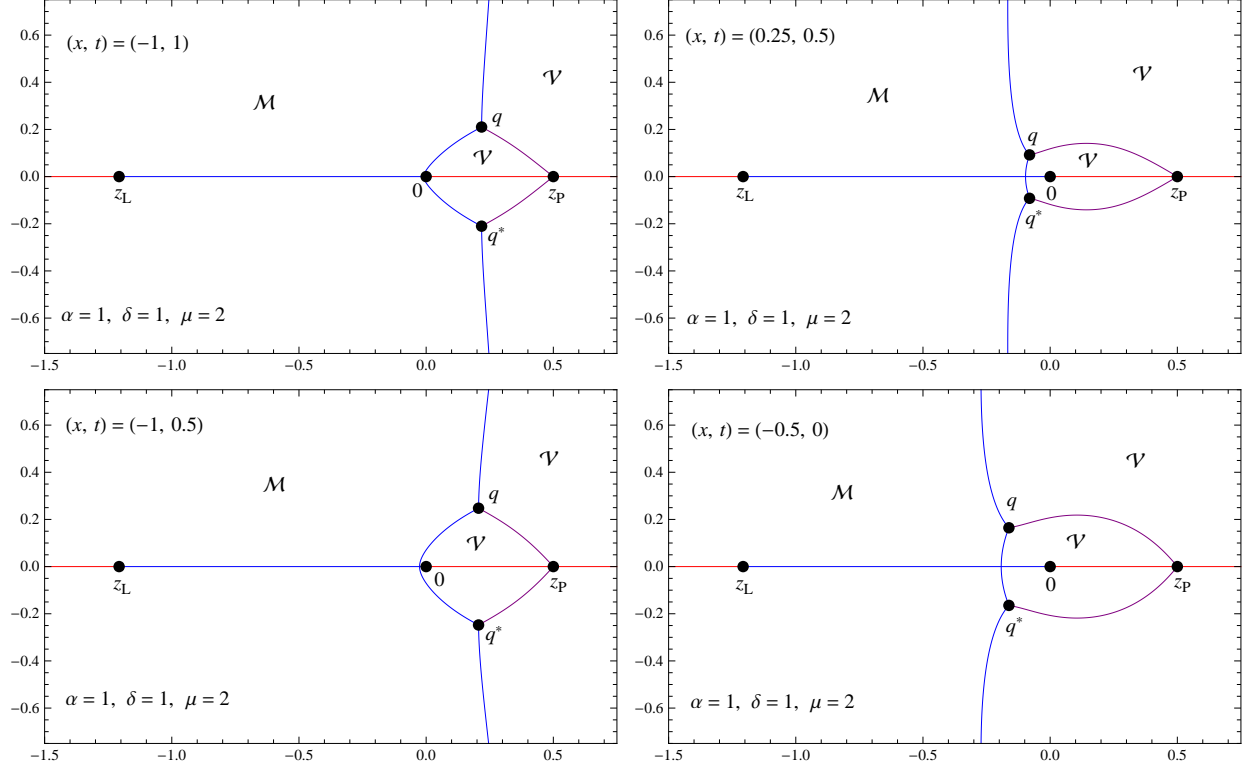


FIGURE 7. Four numerically generated sign charts for $\Im\{h(z)\}$ indicating the regions \mathcal{M} and \mathcal{V} , all for parameter values $\alpha = 1$, $\delta = 1$, and $\mu = 2$. Upper left (UL): $(x, t) = (-1, 1)$. Upper right (UR): $(x, t) = (0.25, 0.5)$. Lower left (LL): $(x, t) = (-1, 0.5)$. Lower right (LR): $(x, t) = (-0.5, 0)$. The red curves are branch cuts for $h(z)$, the blue curves are zero level curves of $\Im\{h(z)\}$, and the purple curve is $\mathcal{B} \cup \mathcal{B}^*$, which is both a branch cut of $h(z)$ (although $\Im\{h(z)\}$ is continuous there) and a zero level curve of $\Im\{h(z)\}$.

Also, the strict inequalities (146)–(147) hold for $\Im\{z\} > 0$, with the only excluded points in the upper half-plane satisfying $\Im\{h(z)\} = 0$. From (142) we have the asymptotic condition

$$(149) \quad \Im\{h(z)\} = -\left[\frac{2\pi}{\alpha^2} + \frac{8t}{\alpha^2}\Im\{z\}\right]\Re\{z\} + O(|\Re\{z\}|^{1/2}), \quad \Re\{z\} \rightarrow -\infty, \quad 0 \leq \Im\{z\} = O(1).$$

Finally, $h(z)$ is continuous at $z = q$, and

$$(150) \quad h'(z)^2 = (z - q)v(z)$$

where v is analytic and non-vanishing in a neighborhood of $z = q$.

Proof. It only remains to prove (148) and (150). But, these both follow from the representation (89). Indeed, taking into account that $S(z)$ changes sign across \mathcal{B} while $Y(z; q, q^*)$ is analytic in a neighborhood of \mathcal{B} , we easily see that the sum of boundary values $h'_+(z) + h'_-(z)$ vanishes identically for $z \in \mathcal{B}$. Therefore, $h_+(z) + h_-(z)$ is constant along \mathcal{B} , and since \mathcal{B} is part of the zero level set \mathcal{Z} of $\Im\{h(z)\}$, it is obvious that the constant value taken by $h_+(z) + h_-(z)$ for $z \in \mathcal{B}$ is real. It then follows by Schwarz symmetry of $h(z)$ that the same identity holds for $z \in \mathcal{B}^*$. This proves (148). The formula (150) also follows directly from (89) using Proposition 1. \square

With the parametric dependence of h on (x, t) completely determined, we can consider the partial derivatives of h with respect to x and t . Recalling the definition (84) of h in terms of g , θ , and Φ , and the fact that $\Phi(z)$ is a function independent of (x, t) , we immediately deduce that these partial derivatives are analytic in

a larger domain than is h itself, namely for $z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*)$. We claim that the partial derivatives are given by the following simple explicit formulae:

$$(151) \quad \frac{\partial h}{\partial x}(z) = \frac{1}{2} \frac{\partial \kappa}{\partial x} - \frac{2}{\alpha} S(z), \quad z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*),$$

and

$$(152) \quad \frac{\partial h}{\partial t}(z) = \frac{1}{2} \frac{\partial \kappa}{\partial t} - \frac{4}{\alpha^2} \left(z + \Re\{q\} - \frac{1}{2} \right) S(z), \quad z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*).$$

Indeed, these formulae exhibit the correct domain of analyticity and capture the correct boundary conditions on $\mathcal{B} \cup \mathcal{B}^*$ following from (148):

$$(153) \quad \frac{\partial h_+}{\partial x}(z) + \frac{\partial h_-}{\partial x}(z) = \frac{\partial \kappa}{\partial x} \quad \text{and} \quad \frac{\partial h_+}{\partial t}(z) + \frac{\partial h_-}{\partial t}(z) = \frac{\partial \kappa}{\partial t}, \quad z \in \mathcal{B} \cup \mathcal{B}^*.$$

Also, since the partial derivatives of g are bounded for $z = \infty$, the principal parts of the partial derivatives of h at $z = \infty$ must agree with those of θ , and by expansion of $S(z)$ for large z it is easy to confirm that the above formulae indeed satisfy

$$(154) \quad \frac{\partial h}{\partial x}(z) = \frac{\partial \theta}{\partial x}(z) + O(1) = -\frac{2}{\alpha} z + O(1) \quad \text{and} \quad \frac{\partial h}{\partial t}(z) = \frac{\partial \theta}{\partial t}(z) + O(1) = -\frac{4}{\alpha^2} z^2 + \frac{2}{\alpha^2} z + O(1), \quad z \rightarrow \infty.$$

Since $g(0) = 0$ for all (x, t) it follows from (84) that the partial derivatives of h must agree with those of θ exactly at $z = 0$. Imposing these conditions on (151) and (152) proves that the partial derivatives of κ can be expressed explicitly in terms of $q(x, t)$. Indeed, since $S(0) = -|q|$,

$$(155) \quad 0 = \frac{\partial h}{\partial x}(0) - \frac{\partial \theta}{\partial x}(0) = \frac{1}{2} \frac{\partial \kappa}{\partial x} + \frac{2}{\alpha} |q| - \frac{1}{2\alpha} \implies \frac{\partial \kappa}{\partial x} = \frac{1}{\alpha} (1 - 4|q|),$$

and

$$(156) \quad 0 = \frac{\partial h}{\partial t}(0) - \frac{\partial \theta}{\partial t}(0) = \frac{1}{2} \frac{\partial \kappa}{\partial t} + \frac{4}{\alpha^2} \left(\Re\{q\} - \frac{1}{2} \right) |q| + \frac{1}{4\alpha^2} \implies \frac{\partial \kappa}{\partial t} = \frac{4}{\alpha^2} (1 - 2\Re\{q\}) |q| - \frac{1}{2\alpha^2}.$$

Similar calculations allow us to establish a direct expression of $\kappa = \kappa(x, t)$ itself. Indeed, since $g(z)$ is analytic for $z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*)$, is well-defined for $z = \infty$, and satisfies $g_+(z) + g_-(z) = 2\theta(z) + \Phi(z) - \kappa$ for $z \in \mathcal{B} \cup \mathcal{B}^*$, it follows that it must be given by

$$(157) \quad \begin{aligned} g(z) &= \frac{S(z)}{2\pi i} \int_{\mathcal{B} \cup \mathcal{B}^*} \frac{2\theta(s) + \Phi(s) - \kappa}{S_+(s)(s-z)} ds \\ &= \theta(z) - \frac{1}{2} \kappa + \frac{2S(z)}{\alpha^2} (\alpha x - t) + \frac{4tS(z)}{\alpha^2} (z + \Re\{q(x, t)\}) + \frac{S(z)}{2\pi i} \int_{\mathcal{B} \cup \mathcal{B}^*} \frac{\Phi(s) ds}{S_+(s)(s-z)}, \quad z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^*). \end{aligned}$$

Enforcing the condition that $g(0) = 0$ and using the fact that $S(0) = -|q(x, t)|$ then gives a formula for $\kappa(x, t)$:

$$(158) \quad \kappa(x, t) = \frac{1}{2\alpha^2} (2\alpha x - t) - \frac{4}{\alpha^2} |q(x, t)| (\alpha x - t) - \frac{8t}{\alpha^2} |q(x, t)| \Re\{q(x, t)\} - \frac{|q(x, t)|}{\pi i} \int_{\mathcal{B} \cup \mathcal{B}^*} \frac{\Phi(s) ds}{s S_+(s)}.$$

In particular, for $t = 0$ we have

$$(159) \quad \kappa(x, 0) = \frac{x}{\alpha} (1 - 4|q(x, 0)|) - \frac{|q(x, 0)|}{\pi i} \int_{\mathcal{B} \cup \mathcal{B}^*} \frac{\Phi(s) ds}{s S_+(s)}, \quad x < x_c,$$

where in the last integral we use $q = q(x, 0)$ and the corresponding arc \mathcal{B} . This formula can be simplified even further by taking (carefully) the limit $x \uparrow x_c$ in which $q(x, 0) \rightarrow z_c < 0$:

$$(160) \quad \lim_{x \uparrow x_c} \kappa(x, 0) = \frac{x_c}{\alpha} (1 + 4z_c) + 2\Phi(0) - \Phi(z_c) + \frac{2z_c}{\pi} \int_0^{z_P} \frac{\Im\{\Phi_+(s)\} ds}{s(s - z_c)}.$$

Note that $\Phi(0)$ is a well-defined real number, and the fact that the integral is convergent at $s = 0$ follows from (53) which shows that $\Im\{\Phi_+(z)\}$ vanishes linearly as $z \downarrow 0$.

5. STEEPEST DESCENT ANALYSIS OF THE RIEMANN-HILBERT PROBLEM

5.1. Opening of Lenses and Introduction of $g(z)$. We begin by making an explicit modification of the matrix $\mathbf{N}(z)$ by “opening lenses” according to the diagram in Figure 8. The most important features that

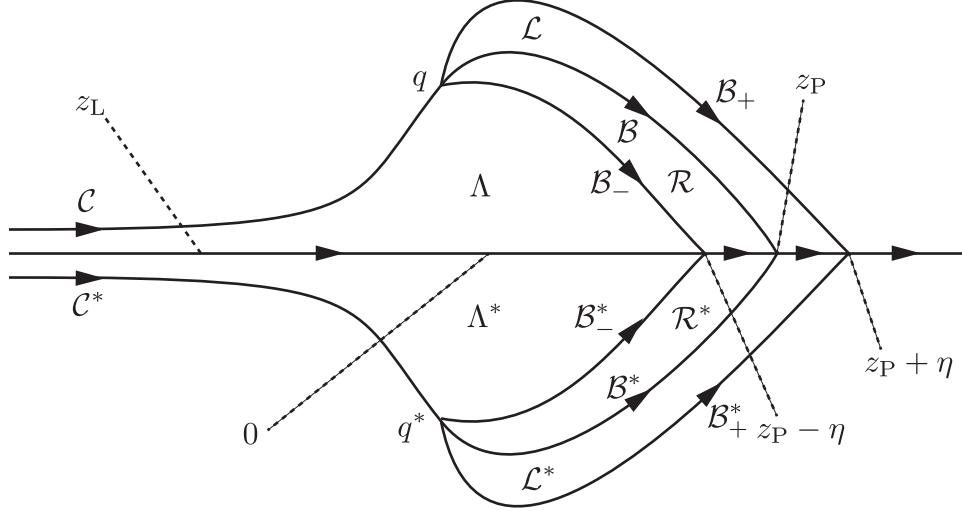


FIGURE 8. The domains Λ , \mathcal{L} , and \mathcal{R} , along with their mirror images in the lower half-plane. Note that $\partial\mathcal{L} \cap \partial\mathcal{R}$ coincides with the contour arc \mathcal{B} . The solid curves comprise a contour denoted Σ .

we require of the contour in this figure are:

- The contour arc \mathcal{C} lies entirely within the mountainous region \mathcal{M} for $\Im\{h(z)\}$, and is asymptotically horizontal as $\Re\{z\} \rightarrow -\infty$.
- The contour arc \mathcal{B}_+ lies entirely within the unbounded component of the valley region \mathcal{V} for $\Im\{h(z)\}$.
- The contour arc \mathcal{B}_- lies entirely within the bounded component of the valley region \mathcal{V} for $\Im\{h(z)\}$.
- The parameter η is sufficiently small that $z_P - \eta > z^+$.

The modification is defined as follows:

$$(161) \quad \mathbf{O}(z) := \mathbf{N}(z) e^{i\theta(z;x,t)\sigma_3/\epsilon} \begin{bmatrix} 1 & -\tilde{E}(z)e^{i\Phi(z)/\epsilon} \\ 0 & 1 \end{bmatrix} e^{-i\theta(z;x,t)\sigma_3/\epsilon}, \quad z \in \Lambda,$$

$$(162) \quad \mathbf{O}(z) := \mathbf{N}(z) e^{i\theta(z;x,t)\sigma_3/\epsilon} \begin{bmatrix} 1 & 0 \\ -\tilde{E}(z)^{-1}e^{-i\Phi(z)/\epsilon} & 1 \end{bmatrix} \tilde{E}(z)^{\sigma_3/2} e^{-i\theta(z;x,t)\sigma_3/\epsilon}, \quad z \in \mathcal{L},$$

$$(163) \quad \mathbf{O}(z) := \mathbf{N}(z) e^{i\theta(z;x,t)\sigma_3/\epsilon} \begin{bmatrix} 0 & -\tilde{E}(z)e^{i\Phi(z)/\epsilon} \\ \tilde{E}(z)^{-1}e^{-i\Phi(z)/\epsilon} & 1 \end{bmatrix} \tilde{E}(z)^{\sigma_3/2} e^{-i\theta(z;x,t)\sigma_3/\epsilon}, \quad z \in \mathcal{R},$$

$$(164) \quad \mathbf{O}(z) := \sigma_1 \mathbf{O}(z^*)^* \sigma_1, \quad z \in \Lambda^* \cup \mathcal{L}^* \cup \mathcal{R}^*,$$

and elsewhere we set $\mathbf{O}(z) := \mathbf{N}(z)$. Recalling that $\tilde{E}(z)$ is an analytic nonvanishing function for z in the open upper half-plane, we see that this defines the matrix $\mathbf{O}(z)$ as an analytic function of z in the complement of the system of contours illustrated in Figure 8. Then, we introduce the function $g(z)$ defined in §4 by setting

$$(165) \quad \mathbf{P}(z) := \mathbf{O}(z) e^{ig(z)\sigma_3/\epsilon}.$$

Since $g(z)$ is analytic for $z \in \mathbb{C} \setminus \mathcal{B} \cup \mathcal{B}^*$, and since $\mathcal{B} = \partial\mathcal{L} \cap \partial\mathcal{R}$, $\mathbf{P}(z)$ has the same domain of analyticity as does $\mathbf{O}(z)$.

The explicit transformation of $\mathbf{N}(z)$ into $\mathbf{P}(z)$ differs from a more standard choice that is based upon the factorization of the jump matrix for $\mathbf{N}(z)$ (say, for $z < 0$) given by

$$(166) \quad \begin{bmatrix} 1 & -s(z)e^{2i\theta(z;x,t)/\epsilon} \\ s(z)^*e^{-2i\theta(z;x,t)/\epsilon} & 1 - |s(z)|^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ s(z)^*e^{-2i\theta(z;x,t)/\epsilon} & 1 \end{bmatrix} \begin{bmatrix} 1 & -s(z)e^{2i\theta(z;x,t)/\epsilon} \\ 0 & 1 \end{bmatrix}.$$

This more standard choice was used, for example, in the analysis of Tovbis, Venakides, and Zhou of the semiclassical limit for the focusing nonlinear Schrödinger equation [21] with special initial data for which there were no discrete eigenvalues (poles of $\mathbf{N}(z)$) and for which the reflection coefficient analogous to $s(z)$ was given in terms of Euler gamma functions. It has the advantage that after opening lenses there remains no jump across the real axis for $z < z_P$. However, opening lenses based on the factorization (166) actually sets into motion a chain of unfortunate events leading to more difficult analysis and even some ambiguity about the scope of the resulting asymptotic formulae. The main problem with using the factorization (166) to open lenses is the presence in the upper half-plane near $z = z_P$ of the *phantom poles* of $s(z)$. These poles both constrain the possible location of the contours near $z = z_P$ (to avoid introducing new singularities into the Riemann-Hilbert problem for $\mathbf{P}(z)$ jump matrices involving $s(z)$ must not be deformed through the phantom poles) and also lead to nonuniformity in the Stirling asymptotics of $s(z)$ near $z = z_P$ that ruins the approximation $E(z) \approx 1$. The authors of [21] dealt with the latter issue by installing an unusual local parametrix near the point analogous to $z = z_P$ in their problem; the parametrix is difficult to analyze because it is not explicit (its existence relies on abstract Fredholm theory) and it depends on ϵ in a way that must be carefully understood. By contrast, the reader will see that the approach we use here (which is based on extending not $s(z)$ but rather the product $(1 + e^{2f(z)/\epsilon})s(z)$ — related to the function $\tilde{E}(z)$ by (51) — into the upper half-plane) *removes all difficulties with the phantom poles near $z = z_P$ and moreover avoids completely the need for any sort of local parametrix near this point.* In fact, $\tilde{E}(z)$ is analytic throughout the upper half-plane, and it will turn out that in our approach the contribution to the error from the neighborhood of $z = z_P$ is exponentially small.

We recall that $g(0) = 0$ and also that $\tilde{E}_+(0) = 0$ (the latter following from the fact that $s(0) = 0$ while $\Phi(0)$ is well-defined and finite) and hence it follows that $\mathbf{P}(z)$ satisfies the normalization condition

$$(167) \quad \mathbf{P}(0) = \mathbb{I}$$

(the value is the same whether the limit is taken from the upper or lower half-plane). Since $g(\infty)$ is well-defined, $\mathbf{P}(z)$ tends to a diagonal limit as $z \rightarrow \infty$ for each fixed ϵ . To describe the jump discontinuities of $\mathbf{P}(z)$ across the arcs of the jump contour pictured in Figure 8, we suppose that all contour arcs are oriented left-to-right (that is, from $-\infty$ toward q , q^* , or $z_P - \eta$, from q or q^* toward $z_P - \eta$, z_P , or $z_P + \eta$, from $z_P - \eta$ toward z_P , from z_P toward $z_P + \eta$, and finally from $z_P + \eta$ toward $+\infty$).

Then the jump conditions satisfied by $\mathbf{P}(z)$ along the real z -axis are as follows:

$$(168) \quad \mathbf{P}_+(z) = \mathbf{P}_-(z) \begin{bmatrix} 1 & e^{2f(z)/\epsilon} s(z) e^{2i(\theta(z;x,t)-g(z))/\epsilon} \\ -e^{2f(z)/\epsilon} s(z)^* e^{-2i(\theta(z;x,t)-g(z))/\epsilon} & 1 - e^{4f(z)/\epsilon} |s(z)|^2 \end{bmatrix}, \quad z < 0,$$

$$(169) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} \begin{bmatrix} 1 & e^{2f(z)/\epsilon} s(z) e^{2i(\theta(z;x,t)-g(z))/\epsilon} \\ e^{2f(z)/\epsilon} s(z)^* e^{-2i(\theta(z;x,t)-g(z))/\epsilon} & 1 + e^{4f(z)/\epsilon} |s(z)|^2 \end{bmatrix}, \quad 0 < z < z_P - \eta,$$

$$(170) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} \begin{bmatrix} A_{11}(z) & A_{12}(z) e^{2i(\theta(z;x,t)-g(z))/\epsilon} \\ A_{21}(z) e^{-2i(\theta(z;x,t)-g(z))/\epsilon} & A_{22}(z) \end{bmatrix}, \quad z_P - \eta < z < z_P,$$

where

$$(171) \quad A_{11}(z) := \frac{1 + |s(z)|^2}{(1 + e^{2f(z)/\epsilon})|s(z)|} e^{\Im\{\Phi_+(z)\}/\epsilon},$$

$$(172) \quad A_{12}(z) = A_{21}(z)^* := \frac{1 - |s(z)|^2 e^{2f(z)/\epsilon}}{(1 + e^{2f(z)/\epsilon})|s(z)|} e^{i\Re\{\Phi_+(z)\}/\epsilon},$$

and where $A_{11}(z)A_{22}(z) - A_{12}(z)A_{21}(z) = 1$,

$$(173) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} \begin{bmatrix} B_{11}(z) & B_{12}(z) e^{2i(\theta(z;x,t)-g(z))/\epsilon} \\ B_{21}(z) e^{-2i(\theta(z;x,t)-g(z))/\epsilon} & B_{22}(z) \end{bmatrix}, \quad z_P < z < z_P + \eta,$$

where

$$(174) \quad B_{22}(z) := \frac{1 + |s(z)|^2}{(1 + e^{2f(z)/\epsilon})|s(z)|} e^{-\Im\{\Phi_+(z)\}/\epsilon},$$

$$(175) \quad B_{12}(z) = B_{21}(z)^* := \frac{|s(z)|^2 e^{2f(z)/\epsilon} - 1}{(1 + e^{2f(z)/\epsilon})|s(z)|} e^{i\Re\{\Phi_+(z)\}/\epsilon},$$

and where $B_{11}(z)B_{22}(z) - B_{12}(z)B_{21}(z) = 1$, and finally

$$(176) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} \begin{bmatrix} 1 & -s(z) e^{2i(\theta(z;x,t)-g(z))/\epsilon} \\ -s(z)^* e^{-2i(\theta(z;x,t)-g(z))/\epsilon} & 1 + |s(z)|^2 \end{bmatrix}, \quad z > z_P + \eta.$$

These jump conditions may appear complicated, and this is the price to be paid for eschewing the standard two-factor factorization (166) in favor of a more complicated one. However, we will now prove that all of the jump conditions for $\mathbf{P}(z)$ along \mathbb{R} amount to exponentially small jump discontinuities, given the condition $\eta < z_P - z^+$. Indeed, using (42) and (43) and recalling that $\theta(z; x, t)$ and $g(z)$ are real for $z \in \mathbb{R}$, we easily see that (168) takes the form

$$(177) \quad \mathbf{P}_+(z) = \mathbf{P}_-(z) \left(\mathbb{I} + O(e^{2f(z)/\epsilon}) \right), \quad z < 0,$$

that (169) takes the form

$$(178) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} \left(\mathbb{I} + O(e^{f(z)/\epsilon}) \right), \quad 0 < z < z_P - \eta,$$

and that (176) takes the form

$$(179) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} \left(\mathbb{I} + O(e^{-f(z)/\epsilon}) \right), \quad z > z_P + \eta.$$

In each case, the jump matrix is a uniformly exponentially small (in ϵ) perturbation of the identity that also decays exponentially as $z \rightarrow \pm\infty$. Also, due to the estimate (44), we have that

$$(180) \quad \mathbf{P}_+(z) = \mathbf{P}_-(z) \left(\mathbb{I} + O(|z|^{1/2} \epsilon^{-1} e^{-M/\epsilon}) \right), \quad -\eta < z < 0,$$

and

$$(181) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} \left(\mathbb{I} + O(|z|^{1/2} \epsilon^{-1} e^{-M/\epsilon}) \right), \quad 0 < z < \eta$$

both hold uniformly for some $\eta > 0$ and $M > 0$. Next, since according to (53) we have $\Im\{\Phi_+(z)\} = -|f(z)|$ for $z > z^+$, we see that for $z_P - \eta < z < z_P$ we have

$$(182) \quad \begin{aligned} A_{11}(z) &= \frac{1 + |s(z)|^2}{(1 + e^{2f(z)/\epsilon})|s(z)|} e^{f(z)/\epsilon} \\ &= 1 + \frac{(|s(z)|e^{f(z)/\epsilon} - 1)(|s(z)|e^{f(z)/\epsilon} - e^{2f(z)/\epsilon})}{|s(z)|e^{f(z)/\epsilon}(1 + e^{2f(z)/\epsilon})} \\ &= 1 + \text{exponentially small in } \epsilon, \quad \text{uniformly for } z_P - \eta < z < z_P, \end{aligned}$$

where we have used (47) and the fact that $e^{2f(z)/\epsilon} \leq 1$ for $z \leq z_P$. Similarly, from (47),

$$(183) \quad \begin{aligned} |A_{12}(z)| = |A_{21}(z)| &= \frac{|1 - |s(z)|^2 e^{2f(z)/\epsilon}|}{2|s(z)|e^{f(z)/\epsilon} \cosh(f(z)/\epsilon)} \\ &\leq \frac{|1 - |s(z)|^2 e^{2f(z)/\epsilon}|}{2|s(z)|e^{f(z)/\epsilon}} \\ &= \text{exponentially small in } \epsilon, \quad \text{uniformly for } z_P - \eta < z < z_P. \end{aligned}$$

Since $f(z)$ changes sign at z_P , we have $e^{-\Im\{\Phi_+(z)\}/\epsilon} = e^{f(z)/\epsilon}$ for $z > z_P$, and therefore

$$\begin{aligned}
(184) \quad B_{22}(z) &= \frac{1 + |s(z)|^2}{(1 + e^{2f(z)/\epsilon})|s(z)|} e^{f(z)/\epsilon} \\
&= 1 + \frac{(|s(z)e^{f(z)/\epsilon} - 1|)(|s(z)e^{f(z)/\epsilon} e^{-2f(z)/\epsilon} - 1|)}{|s(z)|e^{f(z)/\epsilon}(1 + e^{-2f(z)/\epsilon})} \\
&= 1 + \text{exponentially small in } \epsilon, \quad \text{uniformly for } z_P < z < z_P + \eta,
\end{aligned}$$

where we have again used (47) and the fact that $e^{-2f(z)/\epsilon} \leq 1$ for $z \geq z_P$. Similarly,

$$\begin{aligned}
(185) \quad |B_{12}(z)| = |B_{21}(z)| &= \frac{||s(z)|^2 e^{2f(z)/\epsilon} - 1|}{2|s(z)|e^{f(z)/\epsilon} \cosh(f(z)/\epsilon)} \\
&\leq \frac{||s(z)|^2 e^{2f(z)/\epsilon} - 1|}{2|s(z)|e^{f(z)/\epsilon}} \\
&= \text{exponentially small in } \epsilon, \quad \text{uniformly for } z_P < z < z_P + \eta.
\end{aligned}$$

Therefore (170) and (173) can be combined to read

$$(186) \quad \mathbf{P}_+(z) = i^{\sigma_3} \mathbf{P}_-(z) i^{-\sigma_3} (\mathbb{I} + \text{uniformly exponentially small in } \epsilon), \quad |z - z_P| < \eta.$$

Now we describe the jump conditions for $\mathbf{P}(z)$ across the various non-real contours pictured in Figure 8 (it suffices to consider those in the open upper half-plane only due to the symmetry

$$(187) \quad \mathbf{P}(z) = \sigma_1 \mathbf{P}(z^*)^* \sigma_1$$

inherited from (70) and (164) via the fact that $g(z^*) = g(z)^*$). First consider the arc \mathcal{B} separating the regions \mathcal{L} and \mathcal{R} in Figure 8, oriented from $z = q$ toward $z = z_P$. Then recalling the function $h(z)$ defined in terms of $g(z)$ by (84), we have

$$(188) \quad \mathbf{P}_+(z) = \mathbf{P}_-(z) \begin{bmatrix} 0 & e^{i(h_+(z)+h_-(z))/\epsilon} \\ -e^{-i(h_+(z)+h_-(z))/\epsilon} & 0 \end{bmatrix} = \mathbf{P}_-(z) \begin{bmatrix} 0 & e^{i\kappa/\epsilon} \\ -e^{-i\kappa/\epsilon} & 0 \end{bmatrix}, \quad z \in \mathcal{B},$$

where we have used (148). Next, let \mathcal{B}_+ denote the oriented arc in Figure 8 from $z = q$ to $z = z_P + \eta$. Then

$$(189) \quad \mathbf{P}_+(z) = \mathbf{P}_-(z) \begin{bmatrix} 1 & 0 \\ e^{-2ih(z)/\epsilon} & 1 \end{bmatrix} \tilde{E}(z)^{-\sigma_3/2}, \quad z \in \mathcal{B}_+.$$

Similarly, let \mathcal{B}_- denote the oriented arc in Figure 8 from $z = q$ to $z = z_P - \eta$. Then

$$(190) \quad \mathbf{P}_+(z) = \mathbf{P}_-(z) \tilde{E}(z)^{\sigma_3/2} \begin{bmatrix} 1 & 0 \\ e^{-2ih(z)/\epsilon} & 1 \end{bmatrix}, \quad z \in \mathcal{B}_-.$$

Finally, consider the unbounded arc of Σ in the upper half-plane, which we denote by \mathcal{C} . We assume that \mathcal{C} is oriented from $z = -\infty$ toward $z = q$. Then

$$(191) \quad \mathbf{P}_+(z) = \mathbf{P}_-(z) \tilde{E}(z)^{\sigma_3/2} \begin{bmatrix} 1 & e^{2ih(z)/\epsilon} \\ 0 & 1 \end{bmatrix} \tilde{E}(z)^{-\sigma_3/2}, \quad z \in \mathcal{C}.$$

It follows from the inequalities (146)–(147) and the asymptotic estimate (52) that all three of the jump conditions (189)–(191) are of the form $\mathbf{P}_+(z) = \mathbf{P}_-(z)(\mathbb{I} + O(\epsilon))$ uniformly for z on the relevant contours bounded away from the common endpoint $z = q$. In fact, for such z the dominant contribution to the error comes from the matrix factors $\tilde{E}(z)^{\pm\sigma_3/2}$ as $e^{\pm 2ih(z)/\epsilon}$ is in each case exponentially small. Moreover, due to the estimate (149), along the unbounded contour \mathcal{C} , we have that $\mathbf{P}_-(z)^{-1} \mathbf{P}_+(z) - \mathbb{I}$ is exponentially decaying as $z \rightarrow \infty$ for each sufficiently small $\epsilon > 0$.

5.2. Parametrix construction. Let D denote a small open disk of radius independent of ϵ centered at the point $z = q$. We now define an ad-hoc approximation to $\mathbf{P}(z)$, the *global parametrix* $\dot{\mathbf{P}}(z)$:

$$(192) \quad \dot{\mathbf{P}}(z) := \begin{cases} \dot{\mathbf{P}}^{\text{in}}(z), & z \in D \setminus \Sigma \\ \sigma_1 \dot{\mathbf{P}}^{\text{in}}(z^*)^* \sigma_1, & z \in D^* \setminus \Sigma \\ \dot{\mathbf{P}}^{\text{out}}(z), & z \in \mathbb{C} \setminus (\Sigma \cup \bar{D} \cup \bar{D}^*). \end{cases}$$

We call $\dot{\mathbf{P}}^{\text{out}}(z)$ the *outer parametrix* and we call $\dot{\mathbf{P}}^{\text{in}}(z)$ the *inner parametrix*. We emphasize that there is no special parametrix needed in any neighborhood of the point $z = z_P$ from which the phantom poles of $s(z)$ emerge into the upper half-plane.

5.2.1. The outer parametrix. The outer parametrix is easy to write down explicitly:

$$(193) \quad \dot{\mathbf{P}}^{\text{out}}(z) := e^{i\kappa\sigma_3/(2\epsilon)} \mathbf{U} \beta(i(-z)^{1/2}) \sigma_3 \mathbf{U}^\dagger e^{-i\kappa\sigma_3/(2\epsilon)}, \quad \mathbf{U} := \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\pi/4} & e^{-i\pi/4} \\ e^{-i\pi/4} & e^{i\pi/4} \end{bmatrix}, \quad \mathbf{U}^{-1} = \mathbf{U}^\dagger.$$

Here, $i(-z)^{1/2} \in \mathbb{C}_+$ and $\beta(k)$ is the function whose fourth power is given explicitly by

$$(194) \quad \beta(k)^4 := \frac{(k - q^{1/2})(k + q^{1/2})}{(k - q^{*1/2})(k + q^{1/2})}$$

where the principal branch of the square root is meant in all cases, whose branch cuts lie along the arcs of $\pm(\mathcal{B} \cup \mathcal{B}^*)^{1/2}$, and for which the specific branch of the one-fourth power is selected so that $\beta(0) = 1$. It is easy to see that $\beta(k)$ is analytic along the imaginary k -axis, and since $\beta(k)^4 > 0$ for imaginary k , $\beta(k) > 0$ for imaginary k as well, leading to the conclusion that $\beta(\infty) = 1$. This allows us to asymptotically expand $\beta(k)$ for large k , leading to the corresponding expansion of $\dot{\mathbf{P}}^{\text{out}}(z)$:

$$(195) \quad \dot{\mathbf{P}}^{\text{out}}(z) = \mathbb{I} + \begin{bmatrix} 0 & \Im\{q^{1/2}\}e^{i\kappa/\epsilon} \\ -\Im\{q^{1/2}\}e^{-i\kappa/\epsilon} & 0 \end{bmatrix} \frac{1}{i(-z)^{1/2}} + O(z^{-1}), \quad z \rightarrow \infty.$$

It is easy to check that $\dot{\mathbf{P}}^{\text{out}}(z)$ is analytic for $z \in \mathbb{C} \setminus (\mathcal{B} \cup \mathcal{B}^* \cup \mathbb{R}_+)$, and that it is bounded independently of ϵ uniformly for $z \in \mathbb{C} \setminus (D \cup D^*)$ and has determinant one. Also, $\dot{\mathbf{P}}^{\text{out}}(0) = \mathbb{I}$, $\dot{\mathbf{P}}^{\text{out}}(z^*)^* = \sigma_1 \dot{\mathbf{P}}^{\text{out}}(z) \sigma_1$, and

$$(196) \quad \dot{\mathbf{P}}_+^{\text{out}}(z) = \dot{\mathbf{P}}_-^{\text{out}}(z) \begin{bmatrix} 0 & e^{i\kappa/\epsilon} \\ -e^{-i\kappa/\epsilon} & 0 \end{bmatrix}, \quad z \in B,$$

which should be compared with (188), while $\dot{\mathbf{P}}_+^{\text{out}}(z) = i\sigma_3 \dot{\mathbf{P}}_-^{\text{out}}(z) i^{-\sigma_3}$ for $z > 0$, which should be compared with (178), (179), and (186).

5.2.2. The inner parametrix. We now construct the inner parametrix in terms of Airy functions [1], following closely the discussion in section 5.2 of [2]. Note that as a consequence of the continuity of $h(z)$ at $z = q$ and the condition (148) we have $2h(q) = \kappa$. Then, using (150) shows that the equation

$$(197) \quad y^3 = (2ih(z) - i\kappa)^2$$

defines three different univalent functions y of z in a neighborhood of $z = q$ each of which maps $z = q$ to $y = 0$ and which differ by factors of the cube roots of unity; we choose the branch $y = y(z)$ for which the image of $\mathcal{B} \cap D$ is a segment of the negative real y -axis that abuts the origin. We then set $\zeta := \epsilon^{-2/3}y(z)$. Then for $z \in \mathcal{B}_+ \cup \mathcal{B}_- \cup \mathcal{C}$ we have $2ih(z) - i\kappa = -\epsilon\zeta^{3/2}$ where the principal branch of the $3/2$ power is meant. Note that in terms of ζ we may write the outer parametrix in the form

$$(198) \quad \dot{\mathbf{P}}^{\text{out}}(z) = e^{i\kappa\sigma_3/(2\epsilon)} \mathbf{H}(z) \epsilon^{\sigma_3/6} \zeta^{\sigma_3/4} \mathbf{U}^\dagger e^{-i\kappa\sigma_3/(2\epsilon)}, \quad z \in \bar{D},$$

where $\mathbf{H}(z)$ is an ϵ -independent analytic matrix function with determinant one, defined for $z \in \bar{D}$ by

$$(199) \quad \mathbf{H}(z) := \mathbf{U}[y(z)^{-1/4} \beta(i(-z)^{1/2})]^{\sigma_3}, \quad z \in \bar{D}.$$

That $\mathbf{H}(z)$ is analytic follows from the fact that $y(z)^{-1/4} \beta(i(-z)^{1/2})$ extends from $D \setminus \mathcal{B}$ to all of D as a single-valued non-vanishing analytic function. Now for convenience write $\xi := (\frac{3}{4})^{2/3} \zeta$ and define a matrix in the ζ -plane by the formulae:

$$(200) \quad \mathbf{Z}(\zeta) := \sqrt{2\pi} \left(\frac{4}{3}\right)^{\sigma_3/6} \begin{bmatrix} e^{-3\pi i/4} \text{Ai}'(\xi) & e^{11\pi i/12} \text{Ai}'(\xi e^{-2\pi i/3}) \\ e^{-\pi i/4} \text{Ai}(\xi) & e^{\pi i/12} \text{Ai}(\xi e^{-2\pi i/3}) \end{bmatrix} e^{2\xi^{3/2}\sigma_3/3}, \quad 0 < \arg(\zeta) < \frac{2\pi}{3},$$

$$(201) \quad \mathbf{Z}(\zeta) := \sqrt{2\pi} \left(\frac{4}{3}\right)^{\sigma_3/6} \begin{bmatrix} e^{-5\pi i/12} \text{Ai}'(\xi e^{2\pi i/3}) & e^{11\pi i/12} \text{Ai}'(\xi e^{-2\pi i/3}) \\ e^{-7\pi i/12} \text{Ai}(\xi e^{2\pi i/3}) & e^{\pi i/12} \text{Ai}(\xi e^{-2\pi i/3}) \end{bmatrix} e^{2\xi^{3/2}\sigma_3/3}, \quad \frac{2\pi}{3} < \arg(\zeta) < \pi,$$

$$(202) \quad \mathbf{Z}(\zeta) := \sqrt{2\pi} \left(\frac{4}{3}\right)^{\sigma_3/6} \begin{bmatrix} e^{11\pi i/12} \text{Ai}'(\xi e^{-2\pi i/3}) & e^{7\pi i/12} \text{Ai}'(\xi e^{2\pi i/3}) \\ e^{\pi i/12} \text{Ai}(\xi e^{-2\pi i/3}) & e^{5\pi i/12} \text{Ai}(\xi e^{2\pi i/3}) \end{bmatrix} e^{2\xi^{3/2}\sigma_3/3}, \quad -\pi < \arg(\zeta) < -\frac{2\pi}{3},$$

$$(203) \quad \mathbf{Z}(\zeta) := \sqrt{2\pi} \left(\frac{4}{3}\right)^{\sigma_3/6} \begin{bmatrix} e^{-3\pi i/4} \text{Ai}'(\xi) & e^{7\pi i/12} \text{Ai}'(\xi e^{2\pi i/3}) \\ e^{-\pi i/4} \text{Ai}(\xi) & e^{5\pi i/12} \text{Ai}(\xi e^{2\pi i/3}) \end{bmatrix} e^{2\xi^{3/2}\sigma_3/3}, \quad -\frac{2\pi}{3} < \arg(\zeta) < 0.$$

From the well-known asymptotic behavior [1] of $\text{Ai}(\xi)$ and $\text{Ai}'(\xi)$, it follows that

$$(204) \quad \mathbf{Z}(\zeta) \mathbf{U} \zeta^{-\sigma_3/4} = \mathbb{I} + \begin{bmatrix} O(\zeta^{-3/2}) & O(\zeta^{-1}) \\ O(\zeta^{-2}) & O(\zeta^{-3/2}) \end{bmatrix},$$

as $\zeta \rightarrow \infty$ in all directions of the complex plane. We define the inner parametrix by setting

$$(205) \quad \dot{\mathbf{P}}^{\text{in}}(z) := e^{i\kappa\sigma_3/(2\epsilon)} \mathbf{H}(z) \epsilon^{\sigma_3/6} \mathbf{Z}(\zeta) e^{-i\kappa\sigma_3/(2\epsilon)} \mathbf{D}(z), \quad z \in D \setminus \Sigma,$$

where $\mathbf{D}(z)$ is a near-identity diagonal matrix factor given by

$$(206) \quad \mathbf{D}(z) := \begin{cases} \mathbb{I}, & z \in D \cap (\mathcal{L} \cup \mathcal{R}) \\ \tilde{E}(z)^{-\sigma_3/2}, & \text{elsewhere in } D \setminus \Sigma. \end{cases}$$

It is a calculation to confirm that the following jump conditions hold, which should be compared with (188)–(191):

$$(207) \quad \dot{\mathbf{P}}_+^{\text{in}}(z) = \dot{\mathbf{P}}_-^{\text{in}}(z) \begin{bmatrix} 0 & e^{i\kappa/\epsilon} \\ -e^{-i\kappa/\epsilon} & 0 \end{bmatrix}, \quad z \in \mathcal{B} \cap D,$$

$$(208) \quad \dot{\mathbf{P}}_+^{\text{in}}(z) = \dot{\mathbf{P}}_-^{\text{in}}(z) \begin{bmatrix} 1 & 0 \\ e^{-2ih(z)/\epsilon} & 1 \end{bmatrix} \tilde{E}(z)^{-\sigma_3/2}, \quad z \in \mathcal{B}_+ \cap D,$$

$$(209) \quad \dot{\mathbf{P}}_+^{\text{in}}(z) = \dot{\mathbf{P}}_-^{\text{in}}(z) \tilde{E}(z)^{\sigma_3/2} \begin{bmatrix} 1 & 0 \\ e^{-2ih(z)/\epsilon} & 1 \end{bmatrix}, \quad z \in \mathcal{B}_- \cap D,$$

and

$$(210) \quad \dot{\mathbf{P}}_+^{\text{in}}(z) = \dot{\mathbf{P}}_-^{\text{in}}(z) \tilde{E}(z)^{\sigma_3/2} \begin{bmatrix} 1 & e^{2ih(z)/\epsilon} \\ 0 & 1 \end{bmatrix} \tilde{E}(z)^{-\sigma_3/2}, \quad z \in \mathcal{C} \cap D.$$

That is, the inner parametrix satisfies *exactly* the same jump conditions satisfied by $\mathbf{P}(z)$ in a neighborhood D of $z = q$.

5.2.3. Mismatch of the global parametrix across ∂D . It is not difficult to see that the inner and outer parametrices match each other quite well on the boundary of the disk D . Indeed, we have

$$(211) \quad \dot{\mathbf{P}}^{\text{in}}(z) \dot{\mathbf{P}}^{\text{out}}(z)^{-1} = e^{i\kappa\sigma_3/(2\epsilon)} \mathbf{H}(z) \epsilon^{\sigma_3/6} \mathbf{Z}(\zeta) \mathbf{U} \zeta^{-\sigma_3/4} \epsilon^{-\sigma_3/6} \mathbf{H}(z)^{-1} e^{-i\kappa\sigma_3/\epsilon} \cdot \dot{\mathbf{P}}^{\text{out}}(z) \mathbf{D}(z) \dot{\mathbf{P}}^{\text{out}}(z)^{-1}, \\ z \in \partial D.$$

But, since $\mathbf{D}(z) = \mathbb{I} + O(\epsilon)$ while $\dot{\mathbf{P}}^{\text{out}}(z)$ and its inverse are uniformly bounded for $z \in \partial D$, the latter factor is obviously $\mathbb{I} + O(\epsilon)$. Also, since $z \in \partial D$ corresponds to $\zeta^{-1} = O(\epsilon^{2/3})$, from (204) we have that $\epsilon^{\sigma_3/6} \mathbf{Z}(\zeta) \mathbf{U} \zeta^{-\sigma_3/4} \epsilon^{-\sigma_3/6} = \mathbb{I} + O(\epsilon)$ for $z \in \partial D$. It follows that

$$(212) \quad \dot{\mathbf{P}}^{\text{in}}(z) \dot{\mathbf{P}}^{\text{out}}(z)^{-1} = \mathbb{I} + O(\epsilon), \quad \text{uniformly for } z \in \partial D.$$

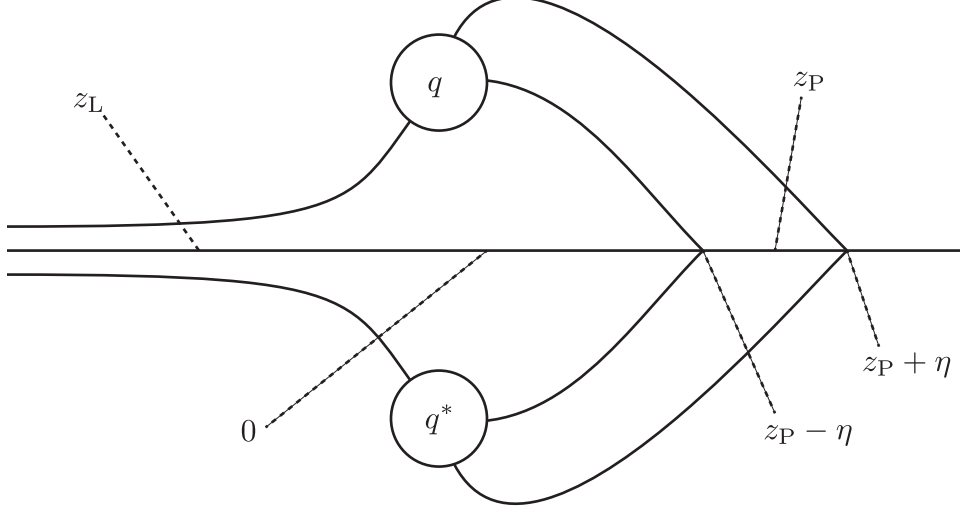


FIGURE 9. The jump contour $\Sigma_{\mathbf{E}}$ for $\mathbf{E}(z)$. The disk boundaries are taken to be oriented in the clockwise direction, and all other contour arcs are oriented left-to-right.

5.3. Error analysis. Let $\mathbf{E}(z) := \mathbf{P}(z)\dot{\mathbf{P}}(z)^{-1}$ denote the mismatch between the (not explicitly known) matrix function $\mathbf{P}(z)$ and its explicit global parametrix $\dot{\mathbf{P}}(z)$. Because $\mathbf{P}(z)$ and $\dot{\mathbf{P}}(z)$ satisfy exactly the same jump conditions on the contours \mathcal{B} , \mathcal{B}^* , and $\Sigma \cap (D \cup D^*)$, $\mathbf{E}(z)$ may be extended analytically to the latter contours. On the other hand, $\mathbf{E}(z)$ has jumps across the disk boundaries ∂D and ∂D^* where $\mathbf{P}(z)$ has no jump. That is, $\mathbf{E}(z)$ may be regarded as an analytic matrix function of $z \in \mathbb{C} \setminus \Sigma_{\mathbf{E}}$, where $\Sigma_{\mathbf{E}}$ is the contour shown in Figure 9. By combining the jump conditions satisfied by $\mathbf{P}(z)$ with those known to be satisfied by the parametrix $\dot{\mathbf{P}}(z)$, it is a simple calculation to see that on the complex arcs of $\Sigma_{\mathbf{E}}$ as well as for $z < 0$, we have that $\mathbf{E}_-(z)^{-1}\mathbf{E}_+(z) - \mathbb{I}$ is $O(\epsilon)$ in both the $L^\infty(\Sigma_{\mathbf{E}} \setminus \mathbb{R}_+)$ and the $L^2(\Sigma_{\mathbf{E}} \setminus \mathbb{R}_+)$ senses. For $z > 0$ we have that $[i^{\sigma_3}\mathbf{E}_-(z)i^{-\sigma_3}]^{-1}\mathbf{E}_+(z) - \mathbb{I}$ is $O(\epsilon)$ (in fact, exponentially small) in both the $L^\infty(\mathbb{R}_+)$ and $L^2(\mathbb{R}_+)$ senses. Because both $\mathbf{P}(0) = \mathbb{I}$ and also $\dot{\mathbf{P}}(0) = \dot{\mathbf{P}}^{\text{out}}(0) = \mathbb{I}$, we have $\mathbf{E}(0) = \mathbb{I}$.

The unusual nature of the jump condition on \mathbb{R}_+ can be understood by introducing the matrix function $\mathbf{F}(w)$ defined by

$$(213) \quad \mathbf{F}(w) = \begin{cases} i^{\sigma_3}\mathbf{E}(w^{-2})i^{-\sigma_3}, & \Im\{w\} > 0 \\ \mathbf{E}(w^{-2}), & \Im\{w\} < 0. \end{cases}$$

The jump contour $\Sigma_{\mathbf{F}}$ for $\mathbf{F}(w)$ consists of reciprocals of the positive and negative square roots of the complex arcs of $\Sigma_{\mathbf{E}}$ along with the real and imaginary w -axes. It follows directly that $\mathbf{V}(w) - \mathbb{I} = O(\epsilon)$ in both the $L^\infty(\Sigma_{\mathbf{F}})$ and $L^2(\Sigma_{\mathbf{F}})$ senses, where $\mathbf{V}(w) := \mathbf{F}_-(w)^{-1}\mathbf{F}_+(w)$ for $w \in \Sigma_{\mathbf{F}}$. Also we unambiguously calculate the value $\mathbf{F}(\infty) = \mathbb{I}$. This problem is solved by writing $\mathbf{F}(w)$ in terms of a Cauchy integral

$$(214) \quad \mathbf{F}(w) = \mathbb{I} + \frac{1}{2\pi i} \int_{\Sigma_{\mathbf{F}}} \frac{(\mathbb{I} + \mathbf{X}(s))(\mathbf{V}(s) - \mathbb{I})}{s - w} ds, \quad w \in \mathbb{C} \setminus \Sigma_{\mathbf{F}},$$

where $\mathbf{X} \in L^2(\Sigma_{\mathbf{F}})$ solves the singular integral equation

$$(215) \quad \mathbf{X}(w) - \frac{1}{2\pi i} \int_{\Sigma_{\mathbf{F}}} \frac{\mathbf{X}(s)(\mathbf{V}(s) - \mathbb{I})}{s - w_-} ds = \frac{1}{2\pi i} \int_{\Sigma_{\mathbf{F}}} \frac{\mathbf{V}(s) - \mathbb{I}}{s - w_-} ds, \quad w \in \Sigma_{\mathbf{F}}.$$

Here the notation w_- indicates that the integrals are to be evaluated for $w \notin \Sigma_{\mathbf{F}}$ and the limit of w approaching a point on $\Sigma_{\mathbf{F}}$ is to be taken from the right side according to the orientation of each arc. By standard theory [23], this equation has a unique solution whose norm is $O(\epsilon)$. Now since $\mathbf{V}(w) - \mathbb{I}$ vanishes

to all orders as $w \rightarrow 0$ along $\Sigma_{\mathbf{F}}$, $\mathbf{F}(w)$ has a nonconvergent but asymptotic power series representation

$$(216) \quad \mathbf{F}(w) - \mathbb{I} \sim \sum_{n=0}^{\infty} \mathbf{F}_n w^n, \quad w \rightarrow 0,$$

with coefficients given by

$$(217) \quad \mathbf{F}_n := \frac{1}{2\pi i} \int_{\Sigma_{\mathbf{F}}} (\mathbb{I} + \mathbf{X}(s))(\mathbf{V}(s) - \mathbb{I}) s^{-n-1} ds, \quad n \geq 0.$$

Now in addition to the estimates $\|\mathbf{X}\|_{L^2(\Sigma_{\mathbf{F}})} = O(\epsilon)$, $\|\mathbf{V} - \mathbb{I}\|_{L^\infty(\Sigma_{\mathbf{F}})} = O(\epsilon)$, and $\|\mathbf{V} - \mathbb{I}\|_{L^2(\Sigma_{\mathbf{F}})} = O(\epsilon)$, we have from (177) and (179) that $\mathbf{V}(w) - \mathbb{I} = O(e^{-M/(\epsilon|w|^2)})$ holds for sufficiently small (independent of ϵ) $|w|$, $w \in \Sigma_{\mathbf{F}}$. Assuming without loss of generality that $|w| < 1$, this implies that $(\mathbf{V}(w) - \mathbb{I})w^{-n-1} = O(\epsilon^{n+1})$ uniformly for small w . These estimates combined with the Cauchy-Schwarz inequality yield that $\mathbf{F}_n = O(\epsilon)$ for all $n \geq 0$. Note that \mathbf{F}_0 is necessarily a diagonal matrix (for consistency with the jump conditions of $\mathbf{E}(z)$ for large z).

From this information, we can now obtain the asymptotic expansion of $\mathbf{P}(z) = \mathbf{E}(z)\dot{\mathbf{P}}(z)$ as $z \rightarrow \infty$: taking $\Im\{w\} < 0$ we have $w = -i(-z)^{-1/2}$, and so $\mathbf{E}(z) = \mathbf{F}(-i(-z)^{-1/2})$. Therefore, since $\dot{\mathbf{P}}(z) = \dot{\mathbf{P}}^{\text{out}}(z)$ for large z , using (195) gives

$$(218) \quad \mathbf{P}(z) = \mathbb{I} + \mathbf{F}_0 + \left((\mathbb{I} + \mathbf{F}_0) \begin{bmatrix} 0 & \Im\{q^{1/2}\}e^{i\kappa/\epsilon} \\ -\Im\{q^{1/2}\}e^{-i\kappa/\epsilon} & 0 \end{bmatrix} - \mathbf{F}_1 \right) \frac{1}{i(-z)^{1/2}} + O(z^{-1}), \quad z \rightarrow \infty.$$

Finally, since for large $z \notin \Lambda \cup \Lambda^*$, we have $\mathbf{M}(i(-z)^{1/2}) = \mathbf{N}(z) = \mathbf{O}(z) = \mathbf{P}(z)e^{-ig(z)\sigma_3/\epsilon}$, we see from (67) that the solution of the MNLS Cauchy problem is

$$(219) \quad \phi(x, t) = \lim_{\substack{z \rightarrow \infty \\ z \notin \Lambda \cup \Lambda^*}} \frac{2i(-z)^{1/2}}{\alpha} \frac{M_{12}(i(-z)^{1/2})}{M_{22}(i(-z)^{1/2})} = \lim_{\substack{z \rightarrow \infty \\ z \notin \Lambda \cup \Lambda^*}} \frac{2i(-z)^{1/2}}{\alpha} \frac{P_{12}(z)}{P_{22}(z)} = \frac{2}{\alpha} \Im\{q^{1/2}\}e^{i\kappa/\epsilon} + O(\epsilon), \quad \epsilon \rightarrow 0.$$

This asymptotic formula has the form written in the statement of Theorem 1 where

$$(220) \quad A(x, t) := \frac{2}{\alpha} \Im\{q(x, t)^{1/2}\} \quad \text{and} \quad S(x, t) := \kappa(x, t).$$

That the derived fields $\rho(x, t) := A(x, t)^2$ and $u(x, t) := S_x(x, t)$ solve the dispersionless MNLS system (9) has already been shown, since according to (155) we can express $u(x, t)$ explicitly in terms of $q(x, t)$, and hence

$$(221) \quad \rho(x, t) = \frac{4}{\alpha^2} \left(\Im\{q(x, t)^{1/2}\} \right)^2 \quad \text{and} \quad u(x, t) = \frac{1}{\alpha} (1 - 4|q(x, t)|).$$

We have already seen (see the discussion at the end of §4.3) that these substitutions reduce the dispersionless MNLS system to diagonal (Riemann invariant) form, reproducing exactly the partial differential equations (133) that were proved in Proposition 5 to be satisfied by $q(x, t)$ and $q(x, t)^*$. Next, we show that $A(x, 0) = A_0(x)$ and that $S(x, 0) = S_0(x)$. To see this, we note that according to Proposition 2, we have $q(x, 0) = \mathfrak{z}(x)$ for $x < x_c$, where $z = \mathfrak{z}(x)$ is the solution in the upper half-plane of the turning point equation (55); this implies that the quadratic $\chi(x; z)$ is a constant multiple of the factored form $(z - \mathfrak{z}(x))(z - \mathfrak{z}(x)^*) = (z - q(x, 0))(z - q(x, 0)^*)$, which in turn implies the identities:

$$(222) \quad \alpha^2 \rho_0(x) + \frac{1}{2}(\alpha u_0(x) - 1) = -[q(x, 0) + q(x, 0)^*] \quad \text{and} \quad \frac{1}{16}(\alpha u_0(x) - 1)^2 = |q(x, 0)|^2.$$

Under the condition that $u_0(x) < 1/\alpha$, it is easy to see that (221) and (222) imply that

$$(223) \quad \rho(x, 0) = \rho_0(x) \quad \text{and} \quad u(x, 0) = u_0(x).$$

But because $\rho_0(x) > 0$, the condition $u_0(x) < 1/\alpha$ follows from the inequality $Q < 0$ that holds for $x < x_c$ at $t = 0$. It is obvious from (220) that $A(x, t) > 0$ for all $x < x_c(t)$ and $t \geq 0$, and hence we can take a positive square root to find that $A(x, 0) = A_0(x) = \text{sech}(x)$. Since we already have $u(x, 0) = u_0(x)$, to prove that $S(x, 0) = S_0(x)$ we simply need to establish the latter identity at some point $x < x_c$. In fact, we examine the limit $x \uparrow x_c$; since according to (11) we have $S_0(x_c) = \delta x_c + \mu \log(\cosh(x_c))$, we simply compare with the explicit expression for the limiting value of $S(x, 0) = \kappa(x, 0)$ as $x \uparrow x_c$ as given by (160); thus the problem is

reduced to checking the equality of two explicit functions of the three real parameters α , δ , and μ over the region constrained by the three inequalities (15)–(17). We have confirmed this equality numerically, and we therefore conclude that indeed $S(x, 0) = S_0(x) = \delta x + \mu \log(\cosh(x))$ holds for all $x < x_c$.

REFERENCES

- [1] Digital Library of Mathematical Functions. *Release date 2012-3-23*. National Institute of Standards and Technology from <http://dlmf.nist.gov/>.
- [2] R. J. Buckingham and P. D. Miller, “The sine-Gordon equation in the semiclassical limit: dynamics of fluxon condensates,” to appear in *Memoirs AMS*. [arXiv:1103.0061](https://arxiv.org/abs/1103.0061).
- [3] T. Claeys and T. Grava, “Universality of the break-up profile for the KdV equation in the small dispersion limit using the Riemann-Hilbert approach,” *Comm. Math. Phys.*, **286**, 979–1009, 2009.
- [4] T. Claeys and T. Grava, “Solitonic asymptotics for the Korteweg-de Vries equation in the small dispersion limit,” *SIAM J. Math. Anal.*, **42**, 2132–2154, 2010.
- [5] T. Claeys and T. Grava, “Painlevé II asymptotics near the leading edge of the oscillatory zone for the Korteweg-de Vries equation in the small-dispersion limit,” *Comm. Pure Appl. Math.*, **63**, 203–232, 2010.
- [6] P. Deift, S. Venakides, and X. Zhou, “New results in small dispersion KdV by an extension of the steepest descent method for Riemann-Hilbert problems,” *Internat. Math. Res. Notices*, **1997**, 286–299, 1997.
- [7] P. Deift and X. Zhou, “A steepest descent method for oscillatory Riemann-Hilbert problems: asymptotics for the mKdV equation”, *Ann. Math.*, **137**, 295–370, 1993.
- [8] J. C. DiFranco and P. D. Miller, “The semiclassical modified nonlinear Schrödinger equation I: Modulation theory and spectral analysis,” *Physica D*, **237**, 947–997, 2008.
- [9] J. C. DiFranco and P. D. Miller, “The semiclassical modified nonlinear Schrödinger equation III: asymptotic analysis of the Cauchy problem. The hyperbolic region for transsonic initial data,” in preparation.
- [10] J. C. DiFranco, P. D. Miller, and B. K. Muir, “On the modified nonlinear Schrödinger equation in the semiclassical limit: supersonic, subsonic, and transsonic behavior,” *Acta Math. Sci.*, **31B**, 2342–2377, 2011.
- [11] E. V. Doktorov, “The modified nonlinear Schrödinger equation: facts and artefacts,” *European Phys. J. B*, **29**, 227–231, 2002.
- [12] E. V. Doktorov, “Dynamics of a subpicosecond dispersion-managed soliton in a fibre: a perturbative analysis,” *J. Mod. Optics*, **53**, 2701–2723, 2006.
- [13] E. V. Doktorov and I. S. Kuten, “The Gordon-Haus effect for modified NLS solitons,” *Europhys. Lett.*, **53**, 22–28, 2001.
- [14] S. Kamvissis, K. D. T.-R. McLaughlin, and P. D. Miller, *Semiclassical soliton ensembles for the focusing nonlinear Schrödinger equation*, Annals of Math. Studies, **154**, Princeton University Press, 2003.
- [15] D. J. Kaup and A. C. Newell, “An exact solution for a derivative nonlinear Schrödinger equation,” *J. Math. Phys.*, **19**, 798–801, 1978.
- [16] A. V. Kitaev and A. H. Vartanian, “Leading-order temporal asymptotics of the modified nonlinear Schrödinger equation: solitonless sector,” *Inverse Prob.*, **13**, 1311–1339, 1997.
- [17] A. V. Kitaev and A. H. Vartanian, “Asymptotics of solutions to the modified nonlinear Schrödinger equation: solitons on a nonvanishing continuous background,” *SIAM J. Math. Anal.*, **30**, 787–832, 1999.
- [18] B. N. Kuvshinov and V. P. Lakhin, “The Riemann invariants and characteristic velocities of Whitham equations for the derivative nonlinear Schrödinger equation,” *Phys. Scr.*, **49**, 257–260, 1994.
- [19] E. Madelung, “Quantum theory in hydrodynamic form,” *Zeitschrift für Physik*, **40**, 322–326, 1926.
- [20] T. Ramond, “Semiclassical study of quantum scattering on the line,” *Comm. Math. Phys.*, **177**, 221–254, 1996.
- [21] A. Tovbis, S. Venakides, and X. Zhou, “On semiclassical (zero dispersion limit) solutions of the focusing nonlinear Schrödinger equation,” *Comm. Pure Appl. Math.*, **57**, 877–985, 2004.
- [22] M. Wadati, K. Konno, and Y. Ichikawa, “Generalization of inverse scattering method,” *J. Phys. Soc. Japan*, **46**, 1965–1966, 1979.
- [23] X. Zhou, “The Riemann-Hilbert problem and inverse scattering,” *SIAM J. Math. Anal.*, **20**, 966–986, 1989.

DEPARTMENT OF MATHEMATICS, SEATTLE UNIVERSITY, 901 12TH AVE., P.O. BOX 222000, SEATTLE, WA 98122
E-mail address: difranco@seattleu.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, EAST HALL, 530 CHURCH ST., ANN ARBOR, MI 48109
E-mail address: millerpd@umich.edu
URL: www.math.lsa.umich.edu/~millerpd